

10/25536

Rec'd PCT/PTO 24 FEB 2005

REC'D 03 OCT 2003

WIPO

PCT

Pl 1070737

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

**UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office**

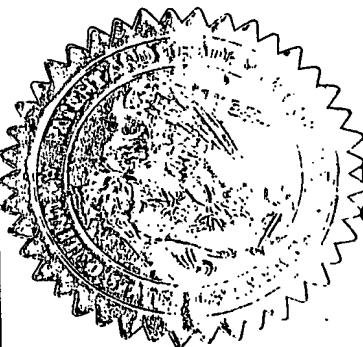
September 26, 2003

**THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.**

APPLICATION NUMBER: 60/406,676

FILING DATE: August 27, 2002

RELATED PCT APPLICATION NUMBER: PCT/US03/26827



**By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS**

W. Montgomery
W. MONTGOMERY
Certifying Officer

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

08/27/02



US31 U.S. PTO

08-29-02

6040246 . 08/27/02

Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

On Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.

EL 701 980 317 US

PTO 0651-0032
08/27/02

08/27/02

INVENTOR(S)					
Given Name (first and middle (if any))	Family Name or Surname	Residence (City and either State or Foreign Country)			
Herve	TETTELIN	Gaithersburg, Maryland, U.S.A.			
<input checked="" type="checkbox"/> Additional inventors are being named on the <u>1</u> separately numbered sheets attached hereto					
TITLE OF THE INVENTION (500 characters max)					
CONSERVED AND SPECIFIC STREPTOCOCCAL GENOMES					
Direct all correspondence to: CORRESPONDENCE ADDRESS					
<input type="checkbox"/> Customer Number		<input type="text"/>		<input type="text"/>	
OR		Type Customer Number here		Place Customer Number Bar Code Label here	
<input checked="" type="checkbox"/> Firm or Individual Name		CHIRON CORPORATION			
Address		P.O. BOX 8097			
Address					
City		Emeryville	State	California	ZIP 94662-8097
Country		U.S.A.	Telephone	510-923-3179	Fax 510-655-3542
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification		Number of Pages <u>104</u>	<input checked="" type="checkbox"/> Includes	<input type="checkbox"/> CD(s), Number <input type="text"/>	
<input checked="" type="checkbox"/> Drawing(s)		Number of Sheets <u>331</u>	<input checked="" type="checkbox"/> Tables	<input type="checkbox"/> Other (specify) <input type="text"/>	
<input type="checkbox"/> Application Data Sheet		See 37 CFR 1.76			
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT					
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.				FILING FEE AMOUNT (\$)	
<input checked="" type="checkbox"/> A check or money order is enclosed to cover the filing fees					
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number:		03-1664		\$160.00	
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.					
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government					
<input checked="" type="checkbox"/> No.					
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: <input type="text"/>					

Respectfully submitted,

SIGNATURE

Rebecca M. Hale

TYPED or PRINTED NAME

Rebecca M. Hale

TELEPHONE

510-923-3179

Date

08/27/2002REGISTRATION NO.
(if appropriate)
Docket Number:

45,680

19195.002

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

504064-012702

CERTIFICATE OF MAILING BY "EXPRESS MAIL" (37 CFR 1.10) Applicant(s): Tettelin, Herve et al.			Docket No. 19195.002
Serial No. To be assigned	Filing Date 08/27/2002	Examiner To be assigned	Group Art Unit To be assigned

Invention: **CONSERVED AND SPECIFIC STREPTOCOCCAL GENOMES**

I hereby certify that the following correspondence:

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

(Identify type of correspondence)

is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 in an envelope addressed to: The Assistant Commissioner for Patents, Washington, D.C. 20231 on

August 27, 2002
(Date)

Rebecca M. Hale
(Typed or Printed Name of Person Mailing Correspondence)

Rebecca M. Hale
(Signature of Person Mailing Correspondence)

EL 701 980 317 US
("Express Mail" Mailing Label Number)

Note: Each paper must have its own certificate of mailing.

PATENT APPLICATION
ATTY REF NO. 19195.002

CONSERVED AND SPECIFIC STREPTOCOCCAL GENOMES

FIELD OF THE INVENTION

The invention relates to polynucleotides which are conserved or specific to one or more species of *Streptococcus*, *Streptococcus* species serotypes, and/or serotype isolates. The conserved or specific genomic regions can be used to identify, screen and develop vaccines and other treatments for Streptococcal infections and can be used in diagnostic assays to diagnose and identify Streptococcal infections.

BACKGROUND OF THE INVENTION

The genus *Streptococcus* consists of Gram-positive, chain-forming, spherical bacterial cells. Three species of clinical interest are *S.pneumoniae* ("pneumococcus" or "S.pn."), *S.pyogenes* ('group A streptococcus' or 'GAS') and *S.agalactiae* ('group B streptococcus' or 'GBS'). Infections with these three pathogenic streptococci lead to conditions including pharyngitis, toxic shock syndrome and necrotizing fasciitis.

Once thought to infect only cows, GBS is now known to cause serious disease, bacteraemia and meningitis in immunocompromised individuals and neonates. There are two known types of neonatal infection. The first (early onset, usually within 5 days of birth) is manifested by bacteraemia and infection. It is generally contracted vertically as a baby passes through the birth canal. GBS is thought to colonize the vagina of about 25% of young women; approximately 1% of infants born via a vaginal birth to colonised mothers will become infected. Mortality resulting from these infections is between 50 – 70%. The second type of neonatal infection is a meningitis that occurs 10 to 60 days after birth. If pregnant women are vaccinated with type III capsule so that the infants are passively immunised, the incidence of the late onset meningitis is generally reduced, although not entirely eliminated.

The "B" in "GBS" refers to the Lancefield classification, which is based on the antigenicity of a carbohydrate which is soluble in dilute acid and called the C carbohydrate. Lancefield identified 13 types of C carbohydrate, designated A to O,

PATENT APPLICATION
ATTY REF NO. 19195.002

that could be serologically differentiated. The organisms that most commonly infect humans are found in groups A, B, D, and G. Within group B, strains can be divided into at least 9 serotypes (Ia, Ib, II, III, IV, V, VI, VII, and VIII) based on the structure of their polysaccharide capsule. Further categories based on, for example, the expression of certain proteins have also been developed.

GBS strains of polysaccharide capsule Type V were rarely isolated before the mid-1980's but now account for approximately one-third of clinical isolates in the US. Type V is the most common capsular serotype associated with invasive infection in nonpregnant adults, and the emergence of Type V strain over the past decade has been temporarily linked to an increase in GBS disease in this population.

Group A streptococcus is a frequent human pathogen, estimated to be present in between 5 – 15% of normal individuals without signs of disease. When host defences are compromised, or when the organism is able to exert its virulence, or when it is introduced into vulnerable tissues or hosts, however, an acute infection occurs. Diseases include puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotising fasciitis, myositis and streptococcal toxic shock syndrome.

Pneumococcus is the most common cause of acute respiratory infection and otitis media and is estimated to result in over 3 million deaths in children every year worldwide from pneumonia, bacteremia, or meningitis. Even more deaths occur among elderly people, among whom *S. pn.* is the leading cause of community-acquired pneumonia and meningitis. Since 1990, the number of penicillin-resistant strains has increased from 1 to 5% to 25 to 80% of isolates, and many strains are now resistant to commonly prescribed antibiotics such as penicillin, macrolides, and fluoroquinolones. See Tettelin, et al. (2001) *Science* 293, 248-506.

The complete genomic sequence of a virulent isolate of *S. pneumoniae* was published by Tettelin, et al. (2001) *Science* 293, 248-506 and is available at the TIGR website at <http://www.tigr.org>. as well as on GEN BANK (available through the Pub Med website at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>). The genomic

PATENT APPLICATION
ATTY REF NO. 19195.002

sequence, the Tettelin article and its published supplemental material are incorporated herein by reference in their entirety.

The complete genomic sequence of an M1 strain of *S. pyrogenes* was published by Ferretti, et al. (2001) *Proc. Natl. Acad. Sci. USA* 98, 4658 – 4663 and is available at the TIGR website at <http://www.tigr.org>. The genomic sequence, the Ferretti article and its published supplemental materials are incorporated herein by reference in their entirety.

The complete genomic sequence of a serotype V strain of *S. agalactiae* (type V strain 2603 V/R) is published on the date of this filing, August 27, 2002 by Gen Bank Accession no. AE009948 (available through Pub Med at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi> and/or is available on the same day at the TIGR website at <http://www.tigr.com>. Most of this sequence is also available in PCT International Patent Application Publication WO 02/34771. The genomic sequence, the Tettelin article and its published supplemental materials are incorporated herein by reference in their entirety.

Current treatments for *Streptococcal* infections include both antibiotics and prophylactic vaccination. Current vaccines, particularly with respect to GBS, suffer from poor immunogenicity, while the emergence of antibiotic resistant strains has lessened the effectiveness of currently used antibiotics. Accordingly, there is an increasing need for the development of new vaccines and antibiotics (as well as other small molecule bacterial inhibitors) to help prevent and treat *Streptococcal* infections.

Applicants have identified regions of the *Streptococcal* genomes which can be used to identify and develop new vaccines and treatments for *Streptococcal* infections. Specifically, Applicants have identified polynucleotides of the *Streptococcal* genome which are conserved or specific to *Streptococcal* species, species serotypes, and/or specific serotype isolates. These polynucleotides and their expressed polypeptides can be used to screen, develop and design new vaccines, antibiotics and other small molecule bacterial inhibitors. These polynucleotides and their expressed polypeptides can further be used to diagnose and identify *Streptococcal* infections.

PATENT APPLICATION
ATTY REF NO. 19195.002

SUMMARY OF THE INVENTION

The invention relates to polynucleotides which are conserved or specific to one or more species of *Streptococcus*, *Streptococcus* species serotypes, and/or serotype isolates. In particular, the invention relates to polynucleotides from *Streptococcus* which are conserved or specific to one or more of the species of *S. pneumoniae* ("pneumococcus" or "S. pn."), *S. pyogenes* ("group A streptococcus" or "GAS"), and *S. agalactiae* ("group B streptococcus" or "GBS"). The invention further relates to polynucleotides which are conserved or specific to one or more Streptococcal species serotypes, such as GBS serotypes Ia, Ib, II, III, IV, V, VI, VII, and VIII. The invention still further relates to polynucleotides which are conserved or specific to one or more clinical isolates of a *Streptococcus* species.

The invention is based on the identification of the following Subsets of genes. Genes falling within each subset are described with respect to referenced tables, lists, and/or figures (in particular the CGH map depicted in Figure 1).

The following Subsets related to the GBS genome:

GBS Subset 1: 1060 GBS genes which have homologs with GAS and with pneumococcus (Table 8);

GBS Subset 2: 225 GBS genes which have homologues with GAS, but not with pneumococcus (Table 10);

GBS Subset 3: 176 GBS genes which have homologues with pneumococcus but not with GAS (Table 9);

GBS Subset 4: 683 GBS genes which do not have homologues with GAS or pneumococcus (specific to GBS vs GAS and pneumococcus) (Table 11).

The invention is based on the identification of the following subsets of genes within the GAS genome:

GAS Subset 1: 1006 GAS genes which have homologues with GBS and with pneumococcus (Table 33);

PATENT APPLICATION
ATTY REF NO. 19195.002

GAS Subset 2: 212 GAS genes which have homologues with GBS but do not have homologues with pneumococcus (Table 34);

GAS Subset 3: 62 GAS genes which have homologues with pneumococcus but do not have homologues with GBS (Table 35);

GAS Subset 4: 416 GAS genes which do not have homologues with either GBS or pneumococcus. This Subset can be determined by subtracting the above subsets from the published genome.

The invention is based on the identification of the following subsets of genes within the pneumococcus genome:

Spn Subset 1: 1034 Spn genes which have homologues with GBS and GAS (Table 36);

Spn Subset 2: 195 Spn genes which have homologues with GBS but do not have homologues with GAS (Table 37);

Spn Subset 3: 74 Spn genes which have homologues with GAS but do not have homologues with GBS (Table 38);

Spn Subset 4: 836 Spn genes which do not have homologues with either GBS or pneumococcus. This Subset can be determined by subtracting the above Subsets from the published genome.

The invention further provides polynucleotides which are conserved or specific to Streptococcus based on a comparison with a wide range of published bacterial genomes. The following additional Subsets are provided:

GBS Subset 1(a): Of the 1060 GBS genes which have homologues in both GAS and pneumococcus, 12 of those GBS genes do not have homologues with any of the other published bacterial genomes at the time of the invention (i.e., GBS Subset 1(a) is specific to Streptococcus vs non Streptococcus published genomes). (The 12 GBS ORF's are listed in Table 3).

GBS Subset 2(a): This Subset comprises GBS genes which have homologues with GAS, but not with pneumococcus or any other published bacterial genomes at the time of the invention.

GBS Subset 3(a): This Subset comprises GBS genes which have homologues with pneumococcus, but not with GAS or any other published bacterial genomes at the time of the invention.

GBS Subset 4(a): Of the 683 GBS genes which do not have homologues in either GAS or pneumococcus, 315 of these GBS genes also do not have homologues with any of the other published bacterial genomes. These include six proteins predicted to be anchored on the cell wall (SAG0677, SAG0771, SAG1052, SAG1331, SAG1473, and SAG1168), three of the capsule-related genes (SAG1163, SAG1167, and SAG1168), six transcriptional regulators, and four genes of the cyl operon (SAG0663 – SAG0673) essential for GBS hemolytic activity and production of pigment. See Pritzlaff et al. (2001) *Mol. Microbiol.*, 39, 236 – 247. The rest of the 315 proteins include 240 hypothetical proteins with no similarity to other proteins in databases.

Many of the 315 genes specific to *S. agalactiae* are located in regions likely to constitute mobile genetic elements. Two of these regions resemble prophages (SAG0545-SAG0610 and SAG1835-SAG1885) displaying a mosaic structure with segments most similar to different bacteriophages, a pattern that suggests frequent recombination events. PblA and PblB are adhesins from a *S. mitis* prophage where they contribute to endocarditis by binding to human platelets (See Bensing, et al. (2001) *Infect. Immun.* **69**, 6186 – 6192; Bensing, et al (2001) *Infect. Immun.* **69**, 1373 – 1380. Their orthologs in *S. agalactiae* are located on separate prophages and display a different protein structure. Another region (SAG1247-SAG1299) encodes a putative conjugative transposon that carries genes for cadmium efflux and mercury resistance.

GAS Subset 1(a): This Subset comprises GAS genes which have homologues with GBS and with pneumococcus, but do not have homologues with any of the other published bacterial genomes at the time of the invention.

PATENT APPLICATION
ATTY REF NO. 19195.002

GAS Subset 2(a): This Subset comprises GAS genes which have homologues with GBS but do not have homologues with pneumococcus or any of the other published bacterial genomes at the time of the invention;

GAS Subset 3(a): This Subset comprises GAS genes which have homologues with pneumococcus but do not have homologues with GBS or any of the other published bacterial genomes at the time of the invention.

GAS Subset 4(a): This Subset comprises GAS genes which do not have homologues with either GBS or pneumococcus or with any of the other published bacterial genomes at the time of the invention.

Spn Subset 1(a): This Subset comprises Spn genes which have homologues with GBS and GAS but which do not have homologues with any of the other published bacterial genomes at the time of the invention;

Spn Subset 2(a): This Subset comprises Spn genes which have homologues with GBS but do not have homologues with GAS or with any of the other published bacterial genomes at the time of the invention;

Spn Subset 3(a): This Subset comprises Spn genes which have homologues with GAS but do not have homologues with GBS or with any of the other published bacterial genomes at the time of the invention;

Spn Subset 4(a): This Subset comprises Spn genes which do not have homologues with either GBS or pneumococcus or with any of the other published bacterial genomes at the time of the invention.

The invention also provides polynucleotides which are conserved or specific to GBS serotypes and/or clinical isolates. Applicants have sequenced 19 GBS genes from a variety of GBS serotypes in 11 different clinical isolates. The sequences of these genes are set forth in Tables 13 - 31. The following additional subsets are provided:

GBS Subset 1(b): of the 1060 GBS genes which have homologues with GAS and with pneumococcus, 47 of these GBS genes vary among the 11 clinical isolates (**GBS Subset 1(b)(i)**). 1013 of these GBS genes are conserved across the 11 clinical

isolates (GBS Subset 1(b)(ii)). These lists can be determined by comparing the genes listed in Table 8 with the Comparative Genome Hybridization in Figure 1.

GBS Subset 2(b): of the 225 GBS genes which have homologues with GAS, but not pneumococcus, 44 of these GBS genes vary among the 11 clinical isolates (GBS Subset 2(b)(i)). 181 of these GBS genes are conserved across the 11 clinical isolates (GBS Subset 2(b)(ii)). These lists can be determined by comparing the genes listed in Table 10 with the Comparative Genome Hybridization in Figure 1.

GBS Subset 3(b): of the 176 GBS genes which have homologues with pneumococcus, 44 of these GBS genes vary among 11 clinical isolates (GBS Subset 3(b)(i)). 132 of these GBS genes are conserved across the 11 clinical isolates (GBS Subset 3(b)(ii)). This list can be determined by comparing the genes listed in Table 9 with the Comparative Genome Hybridization in Figure 1.

GBS Subset 4(b): of the 683 GBS genes which do not have homologues with GAS or pneumococcus, 260 GBS genes vary among the 11 clinical isolates (GBS Subset 4(b)(i)). 423 of these GBS genes are conserved across the 11 clinical isolates (GBS Subset 4(b)(ii)). This list can be determined by comparing the genes listed in Table 11 with the Comparative Genome Hybridization in Figure 1. GBS Subset 4(b)(ii) also includes the GBS ORF's listed on Table 12 receiving a "+" under the column "GBS specific".

The invention further provides polynucleotides which are likely recent genomic duplications in GBS. These duplications include glycosyl transferases, sortases, proteins anchored on the cell wall, β lactam resistance factors, and many hypothetic proteins. The GBS genes are listed in Table 4 (GBS Subset 5).

The invention is also based on the identification of a cluster of 13 adjacent genes (SAG1410 – SAG1424) which is believed to encode enzymes required for synthesis of the group B carbohydrate, a complex multiantennary structure of rhamnose, glucitol phosphate, N-acetylglucosamine, and galactose. (GBS Subset 6). Predicted proteins encoded within this cluster include seven putative glycosyltransferases, four of which are similar to rhamnosyltransferases in other streptococcal species; a

PATENT APPLICATION
ATTY REF NO. 19195.002

putative dTDP-L-rhamnose synthase; and proteins involved in glucitol synthesis. All nine regonized GBS capsular polysaccharide types contain sialic acid residues as part of their repeating unit structure, a feature that contributes to virulence by inhibiting activation of the alternative complement pathway. See Edwards et al. (1982) *J. Immunol.* 128, 1278 – 1283.

The type V capsular polysaccharide gene cluster consists of 18 genes. (GBS Subset 6(a)). A region of glycosyltransferases and related proteins (SAG1162 – SAG1170) that direct the synthesis of the type V polysaccharide repeat unit is flanked on either side by genes that are conserved in all known GBS capsule serotypes. Downstream of this region are genes that encode enzynmes for the biosynthesis and activation of sialic acid (SAG1158 – SAG1161). Upstream of the serotype specific region are genes (SAG1171 – SAG1175) found not only in all nine GBS capsular serotypes but also in a variety of other polysaccharide-producing streptococci.

The invention is also based on the identification of GBS ORFs predicted to encode proteins carrying a signal peptide (GBS Subset 7). These GBS ORF's are listed in Table 2 receiving a "+" under the column "signal peptide".

The invention is also based on the identification of GBS ORFs predicted to encode proteins which are anchored on the cell wall through an LPxTG motif (GBS Subset 8). These GBS ORF's are listed in Table 2 receiving a "+" under the column "sortase motif".

The invention is also based on the identification of GBS ORFs prediced to encode lipoproteins (GBS Subset 9). These GBS ORF's are listed in Table 2 receiving a "+" under the column "lipoprotein".

The invention is also based on the identification of two GBS ORF's predicted to encode enzymes related to metabolism (GBS Subset 10). These GBS ORFs include a putative pullulanase (SAG1216) and a neuraminidase-related protein (SAG1932).

PATENT APPLICATION
ATTY REF NO. 19195.002

The invention is also based on the identification of GBS ORF's predicted to encode proteins exposed on the cell surface (GBS Subset 11). These GBS ORF's are listed in Table 2 receiving a "+" under the column "FACS".

The invention is also based on the identification of 401 GBS ORF's from GBS strain 2603 V/R which were not detected in at least one other of the 11 tested clinical isolates (GBS Subset 12). See Comparative Hybridization Genome in Figure 1. 364 of these 401 ORF's correspond to 15 regions containing more than 5 contiguous genes. Each region is identified in Figure 1 by numerical yellow bullets. Each region comprises a subset as defined below:

Region 1: **GBS Subset 12(a)**. This region is unique to GBS (SAG0218 – SAG0238). This region is a possible plasmid or remnant of a phage and contains mostly hypothetical proteins.

Region 2: **GBS Subset 12(b)**

Region 3: **GBS Subset 12(c)**

Region 4: **GBS Subset 12(d)**

Region 5: **GBS Subset 12(e)**

Region 6: **GBS Subset 12(f)**

Region 7: **GBS Subset 12(g)**

Region 8: **GBS Subset 12(h)**. This region is specific to GBS (SAG1018 – SAG1037). This region comprises 20 proteins of unknown function, most of which are predicted to be membrane associated or secreted, and displays an atypical nucleotide composition.

Region 9: **GBS Subset 12(i)**

Region 10: **GBS Subset 12(j)**

Region 11: **GBS Subset 12(k)**

Region 12: **GBS Subset 12(l)**

Region 13: **GBS Subset 12(m)**

PATENT APPLICATION
ATTY REF NO. 19195.002

Region 14: GBS Subset 12(n). This region is unique to GBS and spans 33 genes (SAG1989 – 2021), including 25 proteins of unknown function, some of which carry a cell-wall anchor.

Region 15: GBS Subset 12(o).

This invention is also based on identification of clusters of GBS genes as set forth in Figure 5 and Table 6. In Figure 5, the presence of a particular gene or gene cluster is indicated in the figure by a red square and the absence of a gene or cluster by a black square. The relationship between strains based on this analysis is depicted by the tree at the top of the figure. The strains and their serotypes are indicated (NT: nontypeable). Clusters with identical profiles are reduced to a single horizontal line and the number of genes in each cluster is indicated on the right. The clusters of 5 or more genes, labeled in red text and numbered, are listed in Table 6. The 1698 genes shared by all 19 strains are labeled in green text. Applicants identified the following subsets:

- GBS Subset 13 (a):** Cluster 1 (from Table 6).
- GBS Subset 13 (b):** Cluster 2 (from Table 6).
- GBS Subset 13 (c):** Cluster 3 (from Table 6).
- GBS Subset 13 (d):** Cluster 4 (from Table 6).
- GBS Subset 13 (e):** Cluster 5 (from Table 6).
- GBS Subset 13 (f):** Cluster 6 (from Table 6).
- GBS Subset 13 (g):** Cluster 7 (from Table 6).
- GBS Subset 13 (h):** Cluster 8 (from Table 6).
- GBS Subset 13 (i):** Cluster 9 (from Table 6).
- GBS Subset 13 (j):** Cluster 10 (from Table 6).
- GBS Subset 13 (k):** Cluster 11 (from Table 6).
- GBS Subset 13 (l):** Cluster 12 (from Table 6).
- GBS Subset 13 (m):** Cluster 13 (from Table 6).
- GBS Subset 13 (n):** Cluster 14 (from Table 6).
- GBS Subset 13 (o):** Cluster 15 (from Table 6).
- GBS Subset 13 (p):** Cluster 16 (from Table 6).
- GBS Subset 13 (q):** 1698 ORFs shared by all strains.

The invention is also based on the identification of the polynucleotide sequences of 19 genes from 11 different GBS strains. The 19 genes are listed on Table 7. A further **GBS Subset 14** includes this set of polynucleotide sequences from the 11 strains and their encoded polypeptide sequences. In particular, GBS Subset 14 contains a Subset of polynucleotide fragments of 10 or more contiguous polynucleotides which are conserved between two or more strains (**GBS Subset 14(a)**). GBS Subset 14 further includes a Subset of polynucleotide fragments of 15 or more contiguous polynucleotides which are conserved between two or more strains (**GBS Subset 14(b)**). GBS Subset 14 further includes a Subset of polynucleotide fragments of 10 or more contiguous polynucleotides which are conserved between three or more strains (**GBS Subset 14(c)**). GBS Subset 14 further includes a Subset of polynucleotide fragments of 10 or more contiguous polynucleotides which are conserved between four or more strains (**GBS Subset 14(d)**).

GBS Subset 14 further includes a Subset of polypeptide fragments of 5 or more contiguous amino acids which are conserved between in two or more strains (**GBS Subset 14(e)**). GBS Subset 14 further includes a Subset of polypeptide fragments of 5 or more contiguous amino acids which are conserved between three or more strains (**GBS Subset 14(f)**). GBS Subset 14 further includes a Subset of polypeptide fragments of 5 or more contiguous amino acids which are conserved between four or more strains (**GBS Subset 14(g)**). GBS Subset 14 further includes a Subset of polypeptide fragments of 10 or more contiguous amino acids which are conserved across two or more strains (**GBS Subset 14(h)**).

The invention provides for methods of screening a Streptococcal genome for a conserved or a specific genomic sequence using one or more of the Subsets of the invention.

The invention further provides for an immunogenic composition comprising a polypeptide expressed by one or more of the polynucleotides in one or more of the Subsets of the invention, and methods for designing an immunogenic composition by

PATENT APPLICATION
ATTY REF NO. 19195.002

selecting one or more polypeptides expressed by one or more of the polynucleotides in one or more of the Subsets of the invention.

The invention further provides for methods of screening compounds for activity against a Streptococcal bacteria, which method comprises contacting the compounds with a polypeptide expressed by the polynucleotide from one of the Subsets of the invention.

The invention further provides for compositions comprising one or more of the polynucleotides, and fragments thereof, selected from the group consisting of the sequences set forth in Tables 13 – 31.

The invention further provides for compositions comprising polypeptides and fragments thereof encoded by the polynucleotides set forth in Tables 13 – 31.

BRIEF DESCRIPTION OF THE TABLES AND DRAWINGS

Table 1 comprises a complete list of GBS predicted genes, listed by SAGxxxx ORF number. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948. This table also includes the predicted amino acid size of the predicted expressed protein and the predicted function, if known.

Table 2 comprises a list of predicted and experimentally characterized surface and secreted proteins from GBS. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 3 lists GBS genes which were shared among GBS, GAS and pneumococcus, but which were not found in any of the other completely sequenced genomes. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 4 depicts GBS genes which are predicted to have been recently duplicated within the genome. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 5 lists the 19 GBS strains used for comparative genome hybridisations and phylogenetic analysis.

Table 6 lists clusters of GBS genes derived from phylogenetic profiling of GBS strains based on comparative genome hybridisations. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 7 lists the GBS genes used for phylogenetic analyses of the 19 GBS strains. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 8 lists the 1060 GBS ORF's which are shared with GAS and pneumococcus. The ORFxxxxx reference number can be translated to SAGxxxx ORF number by using Table 32. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 9 lists the 176 GBS ORF's which are shared with pneumococcus but which are not homologous to a GAS gene. The ORFxxxxx reference number can be translated to SAGxxxx ORF number by using Table 32. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

PATENT APPLICATION
ATTY REF NO. 19195.002

Table 10 lists the 225 GBS ORF's which are shared with GAS but which are not homologous with a pneumococcus gene. The ORFxxxxx reference number can be translated to SAGxxxx ORF number by using Table 32. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 11 lists 683 GBS ORF's which are not shared with either GAS or pneumococcus. The ORFxxxxx reference number can be translated to SAGxxxx ORF number by using Table 32. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 12 lists 315 GBS ORF's which are not shared with GAS, pneumococcus or any other published genomic sequence. The ORFxxxxx reference number can be translated to SAGxxxx ORF number by using Table 32. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 13 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG0466. An alignment of each of the sequences is also included.

Table 14 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG0471. An alignment of each of the sequences is also included.

Table 15 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG0492. An alignment of each of the sequences is also included.

Table 16 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG0767. An alignment of each of the sequences is also included.

Table 17 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG1086. An alignment of each of the sequences is also included.

Table 18 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG1600. An alignment of each of the sequences is also included.

Table 19 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG1680. An alignment of each of the sequences is also included.

Table 20 lists the polynucleotide sequences of the 11 strains relating to GBS ORF SAG1723. An alignment of each of the sequences is also included.

Table 21 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0079. An alignment of each of the sequences is also included.

Table 22 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0093. An alignment of each of the sequences is also included.

Table 23 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0163. An alignment of each of the sequences is also included.

Table 24 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0290. An alignment of each of the sequences is also included.

Table 25 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0368. An alignment of each of the sequences is also included.

Table 26 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG0503. An alignment of each of the sequences is also included.

Table 27 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG1473. An alignment of each of the sequences is also included.

PATENT APPLICATION
ATTY REF NO. 19195.002

Table 28 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG1552. An alignment of each of the sequences is also included.

Table 29 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG1641. An alignment of each of the sequences is also included.

Table 30 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG2147. An alignment of each of the sequences is also included.

Table 31 lists the polynucleotide and polypeptide sequences of the 11 strains relating to GBS ORF SAG2148. An alignment of each of the sequences is also included.

Table 32 provides a conversion table for the ORFxxxx reference numbers to the SAGxxxx reference numbers. The SAGxxxx ORF number corresponds to the genomic sequence for the Streptococcus agalactiae type V strain 2603 V/R available either at the TIGR website on the date of this filing at <http://www.tigr.org> or at the GenBank database at accession number AE009948.

Table 33 lists the 1006 GAS ORF's which are shared with GBS and Spn. The sequences corresponding to these ORFs were published in GenBank, Accession No. AAK33146 (protein sequence). A link to the corresponding polynucleotide sequence is also available. The numbers for the GAS ORF refer directly to their GenBank entries.

Table 34 lists the 212 GAS ORF's which are shared with GBS but which do not have homologues with pneumococcus. The sequences corresponding to these ORFs were published in GenBank, Accession No. AAK33146 (protein sequence). A link to the corresponding polynucleotide sequence is also available. The numbers for the GAS ORF refer directly to their GenBank entries.

Table 35 lists the 62 GAS ORF's which have homologues with pneumococcus but which do not have homologues with GBS. The sequences corresponding to these

ORFs were published in GenBank, Accession No. AAK33146 (protein sequence). A link to the corresponding polynucleotide sequence is also available. The numbers for the GAS ORF refer directly to their GenBank entries.

Table 36 lists the 1034 Spn ORF's which are shared with GBS and GAS. These ORF's were published in GenBank. The numbers for Spn correspond to the entry for AE005672.

Table 37 lists the 195 Spn ORF's which are shared with GBS but do not have homologues with GAS. These ORF's were published in GenBank. The numbers for Spn correspond to the entry for AE005672.

Table 38 lists the 74 Spn ORF's which are shared with GAS but do not have homologues with GBS. These ORF's were published in GenBank. The numbers for Spn correspond to the entry for AE005672.

Figure 1 is a circular representation of the GBS genome and comparative hybridisations using microarrays.

Figure 2 is a schematic representation of in silico comparisons between streptococci.

Figure 3 depicts a phylogenetic tree of GBS strains based on PCR sequences.

Figure 4 depicts a linear representation of the GBS genome.

Figure 5 demonstrates phylogenetic profiling of GBS strains based on comparative genome hybridisations.

BRIEF DESCRIPTION OF THE SEQUENCE ID NOS.

The following SEQ ID NOS are used in the application and figures.

SEQ ID NOS. 1301 – 1316 represent the polynucleotide sequences corresponding to the SAG0466 ORF (thiolase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 1401 – 1417 represent the polynucleotide sequences corresponding to the SAG0471 ORF (glucokinase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

PATENT APPLICATION
ATTY REF NO. 19195.002

SEQ ID NOS. 1501 – 1511 represent the polynucleotide sequences corresponding to the SAG0492 ORF (amino acid ABC transporter, ATP-binding protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 1601 – 1617 represent the polynucleotide sequences corresponding to the SAG0767 ORF (D-alanine - D-alanine ligase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 1701 – 1711 represent the polynucleotide sequences corresponding to the SAG1086 ORF (xanthine phosphoribosyltransferase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 1801 – 1814 represent the polynucleotide sequences corresponding to the SAG1600 ORF (glutamate racemase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 1901 – 1914 represent the polynucleotide sequences corresponding to the SAG1680 ORF (shikimate 5-dehydrogenase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2001 – 2010 represent the polynucleotide sequences corresponding to the SAG1723 ORF (signal peptidase I) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2101 – 2112 represent the polynucleotide sequences corresponding to the SAG0079 ORF (adenylate kinase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2201 – 2211 represent the polynucleotide sequences corresponding to the SAG0093 ORF (D-alanyl-D-alanine carboxypeptidase family protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2301 – 2311 represent the polynucleotide sequences corresponding to the SAG0163 ORF (competence protein Cg1A) in the GBS strains indicated for each sequence, including where indicated reverse complements.

PATENT APPLICATION
ATTY REF NO. 19195.002

SEQ ID NOS. 2401 – 2410 represent the polynucleotide sequences corresponding to the SAG0290 ORF (ABC transporter, substrate-binding protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2501 – 2511 represent the polynucleotide sequences corresponding to the SAG0368 ORF (protein of unknown function) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2601 – 2609 represent the polynucleotide sequences corresponding to the SAG0503 ORF (lipase/acylhydrolase) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2701 – 2711 represent the polynucleotide sequences corresponding to the SAG1473 ORF (cell wall surface anchor family protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2801 – 2811 represent the polynucleotide sequences corresponding to the SAG1552 ORF (conserved hypothetical protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 2901 – 2911 represent the polynucleotide sequences corresponding to the SAG1641 ORF (YaeC family protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 3001 – 3010 represent the polynucleotide sequences corresponding to the SAG2147 ORF (protein of unknown function / lipoprotein, putative) in the GBS strains indicated for each sequence, including where indicated reverse complements.

SEQ ID NOS. 3101 – 3111 represent the polynucleotide sequences corresponding to the SAG2148 ORF (LysM domain protein) in the GBS strains indicated for each sequence, including where indicated reverse complements.

PATENT APPLICATION
ATTY REF NO. 19195.002

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to polynucleotides which are conserved or specific to one or more species of *Streptococcus*, *Streptococcus* species serotypes, and/or serotype isolates. In particular, the invention relates to polynucleotides from *Streptococcus* which are conserved or specific to one or more of the species of *S. pneumoniae* ("pneumococcus" or "S. pn."), *S. pyogenes* ("group A streptococcus" or "GAS"), and *S. agalactiae* ("group B streptococcus" or "GBS"). The invention further relates to polynucleotides which are conserved or specific to one or more Streptococcal species serotypes, such as GBS serotypes Ia, Ib, II, III, IV, V, VI, VII, and VIII. The invention still further relates to polynucleotides which are conserved or specific to one or more clinical isolates of a *Streptococcus* species.

In order to facilitate an understanding of the invention, selected terms used in the application will be discussed below.

As used herein, the phrase "species of *Streptococcus*" generally refers to species of the *Streptococcus* family, including *S. pneumoniae* ("pneumococcus" or "S. pn."), *S. pyogenes* ("group A streptococcus" or "GAS") and *S. agalactiae* ("group B streptococcus" or "GBS").

As used herein, the phrase "*Streptococcus* species serotypes" generally refers to subdivisions based on a distinguishing characteristic within a specific *Streptococcus* species. The distinguishing characteristic can be identified by any of a wide range of diagnostic tools. For instance, GBS is generally recognized as comprising at least nine subdividing serotypes based on the structure of their polysaccharide capsule.

As used herein, the phrases "serotype isolates" or "clinical isolates" generally refer to specific isolated bacterial strains of a specific Streptococcal species and serotype.

As used herein in reference to bacterial genomes, the phrases "conserved" or "shared" generally refer to genomic sequences which have homologues in the two or more genomes in the reference. Homology references, as used in this application, are

PATENT APPLICATION
ATTY REF NO. 19195.002

based on comparisons using FASTA3. *See* Pearson (2000)*Methods Mol. Biol.* 132 185–219. When the homology reference involves a comparison between genes in GBS, GAS or Spn, homologous or shared genes are defined by using a FASTA3 *P* value cutoff of 10^{-15} . Where the homology reference involves a comparison between GBS, GAS or Spn and all other completely sequenced genomes, homologous or shared genes are defined by using a FASTA3 *P* value cutoff of 10^{-5} or lower.

As used herein in reference to bacterial genomes, the phrases "specific to" or "not shared" generally refer to genomic sequences which do not have homologues in the two or more genomes in the reference.

Other software programs to compare identity between nucleotide sequences are known in the art, for example those described in section 7.7.18 of *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30. A preferred alignment program is GCG Gap (Genetics Computer Group, Wisconsin, Suite Version 10.1), preferably using default parameters, which are as follows: open gap = 3; extend gap = 1.

Sequences within a Subset of the invention include sequences which hybridize to the listed genes. Hybridization reactions can be performed under conditions of different "stringency". Conditions that increase stringency of a hybridization reaction of widely known and published in the art [*e.g.* page 7.52 of Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*. NY, Cold Spring Harbor Laboratory]. Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of 25°C, 37°C, 50°C, 55°C and 68°C; buffer concentrations of 10 x SSC, 6 x SSC, 1 x SSC, 0.1 x SSC (where SSC is 0.15 M NaCl and 15 mM citrate buffer) and their equivalents using other buffer systems; formamide concentrations of 0%, 25%, 50%, and 75%; incubation times from 5 minutes to 24 hours; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 x SSC, 1 x SSC, 0.1 x SSC, or de-ionized water. Hybridization techniques and their optimization are well known in the art [*e.g.* see Sambrook *et al.*; *RNA Methodologies* (Farrell, 1998) (Academic Press; ISBN 0-12-249695-7); *Current*

PATENT APPLICATION
ATTY REF NO. 19195.002

Protocols in Molecular Biology (F.M. Ausubel *et al.*, eds., 1987) Supplement 30;
Short protocols in molecular biology (4th edition, 1999) Ausubel *et al.* eds. ISBN 0-471-32938-X; US patent 5,707,829 *etc.*].

Identity between polypeptide sequences can be determined using software programs known in the art, for example those described in section 7.7.18 of *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30. A preferred alignment is determined by the Smith-Waterman homology search algorithm [Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482-489.] using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix 62.

Typically, 50% identity or more between two proteins may be considered to be an indication of functional equivalence. References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences.

The terms "polypeptide", "protein" and "amino acid sequence" as used herein generally refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, mulimers, and the like, are included within the definition. Both full-length proteins and fragments thereof are encompassed by the definition. Minimum fragments of polypeptides useful in the invention can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or even 15 amino acids. Typically, polypeptides useful in this invention can have a maximum length suitable for the intended application. Generally, the maximum length is not critical and can easily be selected by one skilled in the art.

Reference to polypeptides and the like also includes derivatives of the amino acid sequences of the invention. Such derivatives can include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation, and the like. Amino acid derivatives can also include modifications to the native sequence, such as deletions, additions and substitutions (generally conservative in nature), so long as the protein maintains the desired activity. These

PATENT APPLICATION
ATTY REF NO. 19195.002

modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the proteins or errors due to PCR amplification. Furthermore, modifications may be made that have one or more of the following effects: reducing toxicity; facilitating cell processing (*e.g.*, secretion, antigen presentation, etc.); and facilitating presentation to B-cells and/or T-cells.

A "recombinant" protein is a protein which has been prepared by recombinant DNA techniques as described herein. In general, the gene of interest is cloned and then expressed in transformed organisms, as described further below. The host organism expressed the foreign gene to produce the protein under expression conditions. The polypeptides of the invention may be prepared by recombinant means.

The term "polynucleotide", as known in the art, generally refers to a nucleic acid molecule. A "polynucleotide" can include both double- and single-stranded sequences and refers to, but is not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic RNA and DNA sequences from viral (*e.g.* RNA and DNA viruses and retroviruses) or prokaryotic DNA, and especially synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA, and includes modifications such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the nucleic acid molecule encodes a therapeutic or antigenic protein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts that produce the antigens. Modifications of polynucleotides may have any number of effects including, for example, facilitating expression of the polypeptide product in a host cell. The term "polynucleotide" further includes DNA, RNA, DNA/RNA hybrids, DNA and RNA analogues such as those containing modified backbones (with modifications in the sugar and/or phosphates *e.g.* phosphorothioates, phosphoramidites *etc.*), and also peptide nucleic acids (PNA) and any other polymer comprising purine and

pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases *etc.* Nucleic acid according to the invention can be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, probes *etc.*).

A polynucleotide can encode a biologically active (*e.g.*, immunogenic or therapeutic) protein or polypeptide. Depending on the nature of the polypeptide encoded by the polynucleotide, a polynucleotide can include as little as 10 nucleotides, *e.g.*, where the polynucleotide encodes an antigen.

By "isolated" is meant, when referring to a polynucleotide or a polypeptide, that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

"Antibody" as known in the art includes one or more biological moieties that, through chemical or physical means, can bind to or associate with an epitope of a polypeptide of interest. The antibodies of the invention specifically bind to infectious prion conformations. The term "antibody" includes antibodies obtained from both polyclonal and monoclonal preparations, as well as the following: hybrid (chimeric) antibody molecules (see, for example, Winter et al. (1991) *Nature* 349: 293-299; and U.S. Patent No. 4,816,567; F(ab')₂ and F(ab) fragments; F_v molecules (non-covalent heterodimers, see, for example, Inbar et al. (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich et al. (1980) *Biochem* 19:4091-4096); single-chain F_v molecules (sF_v) (see, for example, Huston et al. (1988) *Proc Natl Acad Sci USA* 85:5897-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack et al. (1992) *Biochem* 31:1579-1584; Cumber et al. (1992) *J Immunology* 149B: 120-126); humanized antibody molecules (see, for example, Riechmann et al. (1988) *Nature* 332:323-327; Verhoevan et al. (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional

fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The term "antibody" further includes antibodies obtained through non-conventional processes, such as phage display.

As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited regarding the species or source of the antibody, nor is it intended to be limited by the manner in which it is made. Thus, the term encompasses antibodies obtained from murine hybridomas, as well as human monoclonal antibodies obtained using human rather than murine hybridomas. See, e.g., Cote, et al. *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

An "immunogenic composition" as used herein refers to a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest. The immunogenicity of the composition or the antigenicity of the molecule may be facilitated by the use of an adjuvant.

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Handbook of Surface and Colloidal Chemistry* (Birdi, K.S. ed., CRC Press, 1997); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to*

Biotechniques Series), 2nd ed. (Newton & Graham eds., 1997, Springer Verlag);
Peters and Dalrymple, *Fields Virology* (2d ed), Fields et al. (eds.), B.N. Raven Press,
New York, NY.

It is understood that the antibodies and methods of this invention are not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety.

Vaccines and Immunisation

The invention provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is conserved across one or more species of *Streptococcus*.

The polynucleotide is preferably conserved across one or more species of *Streptococcus* selected from the group consisting of GBS, GAS and pneumococcus. In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous with at least one gene from both GAS and pneumococcus. Preferably, the GBS polynucleotide is selected from GBS Subset 1, which includes 1060 GBS genes which have homologues with both GAS and pneumococcus (Table 8).

In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous with at least one gene from both GBS and pneumococcus. Preferably, the GAS polynucleotide is selected from GAS Subset 1, which includes 1006 GAS genes which have homologues with both GBS and pneumococcus.

In another embodiment, the polynucleotide is a pneumococcal polynucleotide which is homologous with at least one gene both GAS and GBS. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 1, which includes 1034 pneumococcal genes which have homologous with both GBS and GAS.

In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous with at least one gene from GAS. Preferably, the polynucleotide is selected from one of the genes listed GBS Subset 2, which includes 225 GBS genes which have homologues with GAS, but not with pneumococcus.

In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous with at least one gene from pneumococcus. Preferably, the polynucleotide is selected from GBS Subset 3, which includes 176 GBS genes which have homologues with pneumococcus.

In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous with at least one gene from GBS. Preferably, the polynucleotide is selected from GAS Subset 2, which includes 212 GAS genes which have a homologue with GBS.

In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous with at least one gene from pneumococcus. Preferably, the polynucleotide is selected from GAS Subset 3, which includes 62 GAS genes which have a homologue with pneumococcus.

In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one gene from GBS. Preferably, the polynucleotide is selected from Spn Subset 2, which includes 195 Spn genes which have a homologue with GBS.

In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one gene from GAS. Preferably, the polynucleotide is selected from Spn Subset 3, which includes 74 Spn genes which have a homologue with GAS.

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to one or more species of Streptococcus.

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide which is

specific to GBS, GAS and pneumococcus. In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from both GAS and pneumococcus. Preferably, the GBS polynucleotide is selected from GBS Subset 1. In an alternative embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from both GAS and pneumococcus, but which is not homologous to a gene in any other published bacterial genome at the time of the invention. Preferably, the GBS polynucleotide is selected from one of the 12 GBS genes included in GBS Subset 1(a). (Table 3).

In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous to at least one gene in both GBS and pneumococcus. Preferably, the GAS polynucleotide is selected from GAS Subset 1. In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous to at least one gene in both GBS and pneumococcus but which is not homologous to any gene in any other published bacterial genome at the time of the invention. Preferably, the GAS polynucleotide is selected from GAS Subset 1(a).

Alternatively, the polynucleotide is a pneumococcus polynucleotide which is homologous to at least one gene in both GBS and GAS. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 1(a). In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous to at least one gene in both GBS and GAS but which does not have a homologue in any other published bacterial genome at the time of the invention. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 1(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to GBS. In one embodiment, the polynucleotide is a GBS polynucleotide which is not homologue to a gene in either GAS or pneumococcus. Preferably, the GBS polynucleotide is selected from one of the 683 GBS genes included in GBS Subset 4. In a further embodiment, the polynucleotide is a GBS polynucleotide which is not homologous to a gene in either GAS or pneumococcus or

any other published bacterial genome at the time of the invention. Preferably, the GBS polynucleotide is selected from one of the 315 GBS genes in GBS Subset 4(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to GAS. In one embodiment, the polynucleotide is a GAS polynucleotide which is not homologous to a gene in either GBS or pneumococcus. Preferably, the GBS polynucleotide is selected from one of the 416 GAS genes included in GAS Subset 4. In a further embodiment, the polynucleotide is a GAS polynucleotide which does not have a homologue in either GBS or pneumococcus or in any other published bacterial genome at the time of the invention. Preferably, the GAS polynucleotide is selected from GAS Subset 4(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to pneumococcus. In one embodiment, the polynucleotide is a pneumococcus polynucleotide which is not homologous to a gene in either GBS or GAS. Preferably, the pneumococcus polynucleotide is selected from one of the 836 Spn genes included in Spn Subset 4. In a further embodiment, the polynucleotide is a pneumococcus polynucleotide which does not have a homologue in either GBS or GAS or in any other published bacterial genome at the time of the invention. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 4(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to GBS and GAS. In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from GAS but is not homologous to a gene from pneumococcus. Preferably, the GBS polynucleotide is selected from one of the 225 GBS genes included in GBS Subset 2. In another embodiment, the GBS polynucleotide is homologous to at least one gene from GAS but is not homologous to any gene from pneumococcus and does not have a

PATENT APPLICATION
ATTY REF NO. 19195.002

homologue in any other published bacterial genome at the time of the invention.

Preferably, the GBS polynucleotide is selected from GBS Subset 2(a).

In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous to at least one gene from GBS but is not homologous to any gene from pneumococcus. Preferably, the GAS polynucleotide is selected from one of the 212 GAS genes included in GAS Subset 2. In another embodiment, the GAS polynucleotide is homologous to at least one gene from GBS but is not homologous to any gene from pneumococcus and does not have a homologous gene with any other published bacterial genome at the time of the invention. Preferably, the GAS polynucleotide is a selected from GAS Subset 2(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to GBS and pneumococcus. In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from pneumococcus but is not homologous to any gene from GAS. Preferably, the GBS polynucleotide is selected from one of the 176 GBS genes included in GBS Subset 3. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous with at least one gene from pneumococcus but is not homologous with any GAS polynucleotide and does not have a homologous gene in any of the other published bacterial genomes at the time of the invention. Preferably, the GBS polynucleotide is selected from GBS Subset 3(a).

In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one gene from GBS, but is not homologous with any gene from GAS. Preferably, the pneumococcus polynucleotide is selected from one of the 195 Spn genes included in Spn Subset 2. In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one gene from GBS, but is not homologous with any gene from GAS and does not have a homologous gene in any other published bacterial genome at the time of the

invention. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 3(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof which is encoded by a polynucleotide sequence which is specific to GAS and pneumococcus. In one embodiment, the polynucleotide is a GAS polynucleotide which is homologous with at least one gene from pneumococcus but is not homologous with any gene from GBS. Preferably, the GAS polynucleotide is selected from one of the 62 GAS genes included in GAS Subset 3. In another embodiment, the polynucleotide is a GAS polynucleotide which is homologous with at least one gene from pneumococcus but is not homologous with any gene from GBS and is not homologous with any gene of any published bacterial genome at the time of the invention. Preferably, the GAS polynucleotide is selected from GAS Subset 3(a).

In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one GAS polynucleotide, but is not homologous with any GBS gene. Preferably, the pneumococcus polynucleotide is selected from one of the 74 Spn genes included in Spn Subset 3. In another embodiment, the polynucleotide is a pneumococcus polynucleotide which is homologous with at least one gene from GAS, but is not homologous with any gene from GBS or with a gene from any other published bacterial genome at the time of the invention. Preferably, the pneumococcus polynucleotide is selected from Spn Subset 3(a).

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to one or more Streptococcal species serotypes. Preferably, the polynucleotide is specific to a Streptococcal species serotype selected from the Streptococcal species GBS, GAS and pneumococcus. More preferably, the polynucleotide is specific to one or more GBS serotypes selected from the group consisting of GBS serotype Ia, Ib, II, III, IV, V, VI, VII and VIII.

PATENT APPLICATION
ATTY REF NO. 19195.002

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is conserved across one or more Streptococcal species serotypes. Preferably, the polynucleotide is specific to a Streptococcal species serotype selected from the Streptococcal species GBS, GAS and pneumococcus. More preferable, the polynucleotide is conserved across one or more GBS serotypes selected from the group consisting of GBS serotype Ia, Ib, II, III, IV, V, VI, VII and VIII.

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is specific to one or more clinical isolates of a Streptococcal species. Preferably, the polynucleotide is specific to a Streptococcal species clinical isolate selected from the Streptococcal species GBS, GAS and pneumococcus. More preferably, the polynucleotide is specific to one or more GBS clinical isolates selected from the clinical isolates identified in Table 5. Still more preferably, the polynucleotide is specific to one or more GBS clinical isolates having one or more genes selected from the genes listed in Table 7.

In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from both GAS and pneumococcus and which varies among clinical isolates. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from both GAS and pneumococcus and which is homologous with at least one gene from at least one of the clinical isolates identified in Table 5. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from both GAS and pneumococcus and which is homologous with at least one gene from each of the clinical isolates identified in Table 5. Preferably, the polynucleotide is selected from one of the genes listed in Table 7.

In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from GAS and is not homologous to any gene from pneumococcus and which varies among clinical isolates. In another embodiment, the

polynucleotide is a GBS polynucleotide which is homologous to at least one gene from GAS and is not homologous to any gene from pneumococcus and which is homologous to at least one gene from at least one of the clinical isolates identified in Table 5. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from GAS and is not homologous to any gene from pneumococcus and which is homologous to at least one gene from each of the clinical isolates identified in Table 5. Preferably, the polynucleotide is selected from one of the genes listed in Table 7.

In one embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from pneumococcus and is not homologous to any gene from GAS and which varies among clinical isolates. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from pneumococcus and is not homologous to any gene from GAS and which is homologous to at least one gene from at least one of the clinical isolates identified in Table 5. In another embodiment, the polynucleotide is a GBS polynucleotide which is homologous to at least one gene from pneumococcus and is not homologous to any gene from GAS and which is homologous to at least one gene from each of the clinical isolates identified in Table 5. Preferably, the polynucleotide is selected from one of the genes listed in Table 7.

In one embodiment, the polynucleotide is a GBS polynucleotide which is not homologous to any gene from GAS or pneumococcus and which varies among clinical isolates. In another embodiment, the polynucleotide is a GBS polynucleotide which is not homologous to any gene from GAS or pneumococcus and which is homologous to at least one gene from at least one of the clinical isolates identified in Table 5. In another embodiment, the polynucleotide is a GBS polynucleotide which is not homologous to any gene from GAS or pneumococcus and which is homologous to at least one gene from each of the clinical isolates identified in Table 5. Preferably, the polynucleotide is selected from one of the genes listed in Table 7.

The invention further provides an immunogenic composition comprising a polypeptide, or a fragment thereof, which is encoded by a polynucleotide sequence which is conserved across one or more clinical isolates of a Streptococcal species. Preferably, the polynucleotide is conserved across one or more Streptococcal clinical isolates selected from the Streptococcal species GBS, GAS and pneumococcus. More preferably, the polynucleotide is conserved across one or more GBS clinical isolates identified in Table 5. Still more preferably, the polynucleotide is conserved across one or more clinical isolates having one or more genes selected from the genes listed in Table 7.

The invention further provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the Subsets of the invention. Accordingly, the invention provides for an immunogenic composition comprising a polypeptide encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 1, GBS Subset 2, GBS Subset 3, GBS Subset 4, GAS Subset 1, GAS Subset 2, GAS Subset 3, GAS Subset 4, Spn Subset 1, Spn Subset 2, Spn Subset 3, Spn Subset 4, GBS Subset 1(a), GBS Subset 2(a), GBS Subset 3(a), GBS Subset 4(a), GAS Subset 1(a), GAS Subset 2(a), GAS Subset 3(a), GAS Subset 4(a), Spn Subset 1(a), Spn Subset 2(a), Spn Subset 3(a), Spn Subset 4(a), GBS Subset 1(b), GBS Subset 2(b), GBS Subset 3(b), GBS Subset 4(b), GBS Subset 5, GBS Subset 6, GBS Subset 6(a), GBS Subset 7, GBS Subset 8, GBS Subset 9, GBS Subset 10, GBS Subset 11, GBS Subset 12, GBS Subset 12(a), GBS Subset 12(b), GBS Subset 12(c), GBS Subset 12(d), GBS Subset 12(e), GBS Subset 12(f), GBS Subset 12(g), GBS Subset 12(h), GBS Subset 12(i), GBS Subset 12(j), GBS Subset 12(k), GBS Subset 12(l), GBS Subset 12(m), GBS Subset 12(n), GBS Subset 12(o), GBS Subset 13(a), GBS Subset 13(b), GBS Subset 13(c), GBS Subset 13(d), GBS Subset 13(e), GBS Subset 13(f), GBS Subset 13(g), GBS Subset 13(h), GBS Subset 13(i), GBS Subset 13(j), GBS Subset 13(k), GBS Subset 13(l), GBS Subset 13(m), GBS Subset 13(n), GBS Subset 13(o), GBS Subset 13(p), GBS Subset 13(q), GBS Subset 14, GBS Subset 14(a), GBS Subset 14(b), GBS

Subset 14(c), GBS Subset 14(d), GBS Subset 14(e), GBS Subset 14(f), GBS Subset 14(g), and GBS Subset 14(h).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 1, GBS Subset 2, GBS Subset 3, and GBS Subset 4.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GAS Subset 1, GAS Subset 2, GAS Subset 3, and GAS Subset 4.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: Spn Subset 1, Spn Subset 2, Spn Subset 3, and Spn Subset 4.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 1(a), GBS Subset 2(a), GBS Subset 3(a), and GBS Subset 4(a).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GAS Subset 1(a), GAS Subset 2(a), GAS Subset 3(a), and GAS Subset 4(a).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: Spn Subset 1(a), Spn Subset 2(a), Spn Subset 3(a), and Spn Subset 4(a).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or

more of the following Subsets: GBS Subset 1(b), GBS Subset 2(b), GBS Subset 3(b), and GBS Subset 4(b).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from GBS Subset 5.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 6 and GBS Subset 6(a).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 7.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 8.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 9.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 10.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 11.

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 12, GBS Subset 12(a), GBS Subset 12(b), GBS Subset 12(c), GBS Subset 12(d), GBS Subset 12(e), GBS Subset 12(f), GBS Subset 12(g), GBS Subset 12(h), GBS Subset 12(i), GBS Subset 12(j), GBS

PATENT APPLICATION
ATTY REF NO. 19195.002

Subset 12(k), GBS Subset 12(l), GBS Subset 12(m), GBS Subset 12(n), and GBS Subset 12(o).

The invention provides for an immunogenic composition comprising a polypeptide, or a fragment thereof, encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 13(a), GBS Subset 13(b), GBS Subset 13(c), GBS Subset 13(d), GBS Subset 13(e), GBS Subset 13(f), GBS Subset 13(g), GBS Subset 13(h), GBS Subset 13(i), GBS Subset 13(j), GBS Subset 13(k), GBS Subset 13(l), GBS Subset 13(m), GBS Subset 13(n), GBS Subset 13(o), GBS Subset 13(p), GBS Subset 13(q).

The invention provides for an immunogenic composition comprising a polypeptide or a fragment or derivative thereof encoded by a polynucleotide selected from one or more of the following Subsets: GBS Subset 14, GBS Subset 14(a), GBS Subset 14(b), GBS Subset 14(c), GBS Subset 14(d), GBS Subset 14(e), GBS Subset 14(f), GBS Subset 14(g), and GBS Subset 14(h).

The invention further provides a method for designing an immunogenic composition, such as a vaccine, by selecting a polypeptide encoded by a polynucleotide selected from one or more of the Subsets of the invention.

The invention provides a method for raising an immune response in a patient by administering any one of the immunogenic compositions set forth above. The choice of immunogenic composition means that the immune response may be reactive against all three of GAS, GBS and streptococcus, may be reactive against only two of the three, or may be reactive only against GBS.

Each of the immunogenic compositions described above may be prepared and administered instead as a polynucleotide where the polypeptide is expressed *in vivo*.

The immune response is preferably an antibody response. It may be a protective immune response. The patient is preferably a human.

PATENT APPLICATION
ATTY REF NO. 19195.002

Essential genes and knockouts

The invention provides a *Streptococcus* bacterium wherein one or more genes within any of the Subsets of this invention have been knocked out. The choice of Subset means that the knocked out gene may be, for instance, a gene found in GBS but not in GAS or pneumococcus (*e.g.* which is involved in the pathogenesis of GBS, but not in the pathogenesis of GAS or pneumococcus, such as binding GBS cellular targets).

Techniques for producing knockout bacteria are well known, and knockout *Streptococci* of various species have been reported [*e.g.* Margolis *et al.* (2001) *Antimicrob. Agents Chemother.* 45:2432-2435; Zhang *et al.* (2000) *Cell* 102:827-837; Nizet *et al.* (2000) *Infect. Immun.* 68:4245-4254; Nizet *et al.* (1997) *Adv. Exp. Med. Biol.* 418:627-630; *etc.*].

The knockout mutation may be situated in the coding region of the gene or may lie within its transcriptional control regions (*e.g.* within its promoter).

The knockout mutation will reduce the level of mRNA encoding the corresponding polypeptide to <1% of that produced by the wild-type bacterium, preferably <0.5%, more preferably <0.1%, and most preferably to 0%.

The knockout mutants of the invention may be used as immunogenic compositions (*e.g.* as vaccines) to prevent streptococcal infection. Such a vaccine may include the mutant as a live attenuated bacterium.

The knockout mutants of the invention may be used to determine whether genes are essential for bacterial survival, either under normal or stress conditions.

Antisense

The invention provides a single-stranded nucleic acid comprising a fragment of x_1 or more nucleotides from a nucleotide sequence selected from one of the Subsets of the invention. The choice of group means that the nucleic acid may be complementary to a gene sequence found in GBS, GAS and pneumococcus, or a gene sequence specific to GBS.

The single-stranded nucleic acid is at least x_1 nucleotides long. The value of x_1 is at least 7 (e.g. 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 etc.). The single-stranded nucleic acid may be at most x_2 nucleotides long, wherein x_2 is 100 or less (e.g. 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 68, 67, 66, 65, 64, 63, 62, 61, 60).

The nucleic acid is preferably of the formula $5'-(N)_a-(X)-(N)_b-3'$, wherein $0 \leq a \leq 15$, $0 \leq b \leq 15$, N is any nucleotide, and X is the fragment as defined above. The values of a and b may independently be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. Each individual nucleotide N in the $-(N)_a-$ and $-(N)_b-$ portions of the nucleic acid may be the same or different. The length of the nucleic acid (i.e. $a+b+x_1$) is preferably x_2 or less.

Antisense inhibition of streptococcal gene expression is known e.g. Sato *et al.* (1998) *FEMS Microbiol Lett* 159:241-245. Antibacterial antisense techniques are also disclosed in international patent applications WO99/02673 and WO99/13893.

The single-stranded nucleic acid may reduce the level of polypeptide expression from the complementary gene to <1% of that produced by the wild-type bacterium, preferably <0.5%, more preferably <0.1%, and most preferably to 0%.

Antisense experiments may be used to determine whether genes are essential for bacterial survival, either under normal or stress conditions.

Screening methods

The invention provides a method for screening compounds, wherein the method involves contacting the compounds with a polypeptide expressed by one or more of the polynucleotides selected from one of the Subsets of the invention. The method may be for screening for agonists of the polypeptides, antagonists, antibiotics etc. The choice of group means, for instance, that the method may be used for

PATENT APPLICATION
ATTY REF NO. 19195.002

identifying an antibiotic with broad anti-streptococcal activity could be identified, or for identifying an antibiotic specific to GBS.

Potential compounds for screening include small organic molecules, peptides, peptoids, polypeptides, lipids, metals, nucleotides, nucleosides, aptamers, polyamines, antibodies, and derivatives thereof. Small organic molecules have a molecular weight between 50 and about 2,500 daltons, and most preferably in the range 200-800 daltons. Complex mixtures of substances, such as extracts containing natural products, compound libraries or the products of mixed combinatorial syntheses also contain potential antagonists.

Typically, a polypeptide is incubated with a test compound, and the mixture is then tested to see if the polypeptide and test compound interact, or to see if the polypeptide's activity is inhibited.

For preferred high-throughput screening methods, all the biochemical steps for this assay are performed in a single solution in, for instance, a test tube or microtitre plate, and the test compounds are analysed initially at a single compound concentration. For the purposes of high throughput screening, the experimental conditions are adjusted to achieve a proportion of test compounds identified as "positive" compounds from amongst the total compounds screened.

The invention also provides a compound identified using these methods. These can be used to treat or prevent streptococcal infection. The compound preferably has an affinity for the adhesion-specific protein of at least 10^{-7} M *e.g.* 10^{-8} M, 10^{-9} M, 10^{-10} M or tighter.

Distinguishing Streptococcal species

The invention provides a method for determining whether a *Streptococcus* bacterium of interest is or is not in the species *agalactiae*, *pyogenes* or *pneumoniae*, comprising the step(s) of: (a) contacting the bacterium with a nucleic acid probe comprising the sequence of a gene selected from one of the Subsets of the invention; and/or (b) contacting the bacterium with an antibody which binds to a polypeptide

encoded by one or more of the polynucleotides of one or more of the Subsets of the invention. The choice of group means, for instance, that the method may be used for distinguishing GBS from GAS and from pneumococcus, or for confirming that a bacterium is not a GAS or pneumococcus.

The method will typically include the further step of detecting the presence or absence of an interaction between the bacterium of interest and the nucleic acid or protein.

The bacterium of interest may be in a cell culture, for example, or may be within a biological sample believed or known to contain a streptococcus. It may be intact or may be, for instance, lysed.

The term "biological sample" encompasses a variety of sample types obtained from an organism and can be used in a diagnostic or monitoring assay. The term encompasses blood and other liquid samples of biological origin, solid tissue samples, such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. The term encompasses samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components. The term encompasses a clinical sample, and also includes cells in cell culture, cell supernatants, cell lysates, serum, plasma, biological fluids, and tissue samples.

GBS 2603 Type V Genomic Sequence

Applicants have sequenced the complete genome sequence of GBS clinical type V isolate 2603 V/R and performed comparative analyses comparing this sequence with other GBS strains, with other species of pathogenic Streptococci and with other known bacterial species. The entire genomic sequence is available as of the filing date of this application at <http://www.tigr.org>. This genomic sequence is incorporated herein by reference in its entirety. The genomic sequence of GBS type V isolate 2603 V/R is also set forth in International Patent Application WO 02/34771.

In one embodiment, the invention relates to the polynucleotides, and fragments and derivatives thereof, set forth in the GBS clinical type V isolate 2603 published genome which are not disclosed within WO 02/34771. The invention further relates to polypeptides expressed by the polynucleotides of the invention.

Applicants have predicted that the GBS 2603 isolate contains approximately 2,176 predicted genes. Each predicted gene is set forth in Table 1, listed by a SAGxxxx ORF number. Table 1 also includes the predicted amino acid size of the predicted expressed protein and the predicted function, if known. The sequence of each SAG reference can be obtained at the TIGR website.

Figure 1 is a circular representation of the GBS genome and comparative hybridisations using microarrays. The outer circle represents predicted coding regions on the plus strand color coded by role categories: violet indicating amino acid biosynthesis; light blue indicating biosynthesis of cofactors, prosthetic groups, and carriers; light green indicating cell envelope; red indicating cellular processes; brown indicating central intermediary metabolism; yellow indicating DNA metabolism; light gray indicating energy metabolism; magenta indicating fatty acid and phospholipid metabolism; pink indicating protein synthesis and fate; orange indicating purines, pyrimidines, nucleosides, and nucleotides; olive indicating regulatory functions and signal transduction; dark green indicating transcription; teal indicating transport and binding proteins; gray indicating unknown function; salmon indicating other categories; blue indicating hypothetical proteins.

The second circle represents predicted coding regions on the minus strand. In the third circle, black represents atypical nucleotide composition curve; green represents most atypical regions; magenta represents insertion elements; red diamonds indicate rRNAs.

Circles 4 – 22 represent comparative hybridisations of strain 2603 V/R with 19 GBS strains. Cy3/Cy5 (2603 V/R signal/test strain) ratio cutoffs were defined arbitrarily as Cy3/Cy5 – 1.0 – 3.0, the gene was present in the test strain, no color was

PATENT APPLICATION
ATTY REF NO. 19195.002

added; Cy3/Cy5 = 3.0 – 10.0, ambiguous result (blue); Cy3/Cy5 > 10, gene absent in test strain (red).

Circles 4 – 9 represent type 1a strains 090, 515, A909, Davis, and DK8. Circles 10 – 11 represent type 1b strains S7 7357b and H36B. Circles 12 – 13 represent type II strains 18RS21 and DK21. Circles 14 – 18 represent type III COH1, COH31, D136C, M732 and M781. Circle 19 represents type V strain CJB111. Circles 20 – 21 represent type VIII strains SMU014 and JM9130013. Circle 22 represents nontypable (NT) strain CJB110. Throughout Figure 1, varying regions of five or more consecutive genes are indicated by yellow bullets.

Figure 4 depicts a linear representation of the GBS genome. The location of predicted coding regions color-coded by biological role (see Figure 1) is displayed. Arrowed boxes represent the direction of transcription for each ORF. The number of membrane-spanning domains predicted by TopPred is displayed as lipid bi-layers on top of ORFs, only for those whose products have five or more predicted membrane spanning regions. Genes coding for rRNAs (16S, 23S, 5S) and tRNAs (clover leaf structure with number of genes) are indicated. Predicted Rho-independent transcriptional terminators are represented by hairpins.

ORF's were predicted by GLIMMER (See, Delcher, et al., (1999) *Nucleic Acids Res.* 27, 4636 – 4641 and Salzberg, et al., (1998) *Nucleic Acids Res.* 26, 544-548) trained with ORFs larger than 600 base pairs from the genomic sequence and GBS genes available in GenBank. All predicted proteins larger than 30 amino acids were searched against a nonredundant protein database. (See Fleischmann, et al., (1995) *Science* 269, 496 - 512). Frame-shifts and point mutations were detected and corrected where appropriate; those remaining were annotated as "authentic frame-shift" or "authentic point mutation". Protein membrane-spanning domains were identified by TOPPRED (See Claros, et al., (1994) *Comput. Appl. Biosci.* 10, 685 - 686). Candidate lipoprotein signal peptides (See Hayashi et al., (1990) *J. Bioenerg. Biomembr.* 22, 451 - 471) were flagged by N-terminal exact matches to the pattern {DERK} (6)-[LIVMFWSTAG] (2)-[LIVMFYSTAGCQ] – [AGS] – C. Putative

PATENT APPLICATION
ATTY REF NO. 19195.002

signal peptides were identified by using SIGNALP (Nielsen, et al., (1997) *Protein Eng.* 10, 1 - 6). Two sets of hidden Markov models were used to determine ORF membership in families and superfamilies: PFAM Ver. 5.5 (Bateman, et al., (2000) *Nucleic Acids Res.* 28, 263 - 266) and TIGRFAMS 1.0 (Haft et al., (2001) *Nucleic Acids Res.* 29, 41 - 43). Domain-based paralogous families were built by performing all-versus-all searches on the protein sequences by using a modified version of a previously described method. (Niermann, et al., (2001) *Proc. Natl. Acad. Sci. USA* 98, 4136 - 4141) Potential lineage-specific gene duplications were estimated by identification of OFRs more similar to ORFs within the GBS genome than to ORFs from other complete genomes. All ORFs were searched with FASTA3 (Pearson (2000) *Methods Mol. Biol.* 132, 185 - 219) against all ORF's from the complete genomes and matches with a FASTA *P* value of 10^{-15} were considered significant.

The genome consists of a circular chromosome of 2,160,266 base pairs with a G+C content of 35.7%. Base pair one of the chromosome was assigned within the putative origin of replication. The genome contains 80 tRNAs, 7rRNAs, and 3 sRNAs. Approximately 78% of the 2,176 predicted genes are transcribed in the same direction as that of DNA replication, a feature also observed in *S. pn.* and other low-GC Gram positive organisms.

Biological roles were assigned to 1,409 (65%) of the genome according to a classification scheme adapted from Riley (1993) *Microbiol. Rev.* 57, 862 - 952. Another 527 predicted proteins (24%) matched proteins of unknown function, and the remaining 240 (11%) had no database match. The expression of 50 of these hypothetical proteins was confirmed by Western Blot analysis, and the proteins were annotated as "proteins of unknown function." A total of 339 paralogous protein families were identified in strain 2603, containing 941 predicted proteins (43% of the total).

The Western Blot analysis was conducted as follows. GBS strain 2603 V/R cells were grown in Todd-Hewitt broth (Difco) to OD600nm = 0.5. The culture was centrifuged for 20 minutes at 5,000 rpm. The supernatant was discarded, and bacteria

PATENT APPLICATION
ATTY REF NO. 19195.002

were washed once with PBS, resuspended in 2 ml of 50 mM Tris-HCl pH 6.8, containing 400 units of Mutanolysin (Sigma), and incubated 2 hours at 37°C. After three cycles of freeze and thaw, cellular debris was removed by centrifugation at 14,000 rpm for 10 minutes, and the protein concentration of the supernatant was measured by the Bio-Rad Protein assay, with BSA as a standard. Purified recombinant proteins (50 ng) and total cell extracts (25 µg) derived from GBS serotype V 2603 V/R strain were separated by SDS/PAGE and electroblotted onto nitrocellulose membranes for 1 hour at 100 V. The membranes were saturated by overnight incubation at 4° C in 5% skimmed milk and 0.1% Tween 20 in PBS and incubated for 1 hour at room temperature with sera from immunized mice diluted 1:500 - 1:1,000 in saturation buffer. To reduce background due to antibodies raised against contaminating *E. coli* proteins, sera were preincubated with *E. coli* protein extracts absorbed on nitrocellulose strips. The membranes were washed twice in 3% skimmed milk and 0.1% Tween 20 in PBS and incubated for 1 hour with a 1:1,000 dilution of horseradish peroxidase-conjugated antimouse Ig (DAKO). After washing with 0.1% Tween 20 in PBS, the membranes were developed with the Opti-4CN Substrate Kit (Bio-Rad).

Table 2 comprises a list of predicted and experimentally characterized surface and secreted proteins from GBS. Candidate signal peptides and lipoprotein motifs were predicted with PSORT [Nakai, K. & Horton, P. (1999) *Trends Biochem Sci* 24, 34-6] and other methods (see methods), sortase motifs (LPxTG) were detected using the FINDPATTERNS program of the GCG Package [Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res* 12, 387-95] and hidden Markov models. Column "Other" indicates proteins carrying other motifs (*e.g.* integrin-binding motif RGD) or are similar to characterized surface-exposed proteins. Western blot results were considered positive when the antibodies revealed a predominant band of the expected molecular weight on the total protein extracts of *S. agalactiae* strain 2603 V/R, ORFs without + or - in this column were not tested in western blot. FACS

analyses were performed for western blot positive proteins only. Western blot and FACS data are displayed only for proteins carrying at least one of the other motifs shown in the table. Column "GBS specific" indicates genes unique to *S. agalactiae* (when compared to other completely sequenced genomes) that are present in all the *S. agalactiae* strains tested in comparative genome hybridization analyses. Finally, only proteins carrying less than 3 predicted transmembrane domains are shown in the table, other proteins are likely to be embedded in the cytoplasmic membrane and are probably not exposed on the organism's surface.

FACS data was collected as follows: GBS 2603 V/R strain cells were grown in Todd-Hewitt broth (Difco) to OD_{600nm} = 0.5. The culture was centrifuged for 20 minutes at 5,000 rpm, and bacteria were washed once with PBS, resuspended in PBS containing 0.05% paraformaldehyde, and incubated for 1 hour at 37°C and then overnight at 4°C. Fifty microliters of fixed bacteria (OD_{600nm} 0.1) was washed once with PBS, resuspended in 20 µl of newborn calf serum (Sigma), and incubated for 1 hour at 4°C in 100µl of preimmune or immune sera and diluted 1:200 in dilution buffer (PBS, 20% newborn calf serum, 0.1% BSA). After centrifugation and washing with 200µl of washing buffer (0.1% BSA in PBS), samples were incubated for 1 hour at 4°C with 50 µl of R-phycoerythrin-conjugated F(ab)₂ goat anti-mouse IgG (Jackson ImmunoResearch) diluted 1:100 in dilution buffer. Cells were washed with 200 µl of washing buffer and resuspended in 200 µl of PBS. Samples were analysed by using a FACS calibur apparatus (Becton Dickinson), and data were analyzed by using CELL QUEST (Becton Dickinson). A shift in mean fluorescence intensity of >75 channels compared with preimmune sera from the same mice was considered positive. This cutoff was determined from the mean plus two standard deviations of shifts obtained with control sera raised against mock purified recombinant proteins from cultures of *E. coli* carrying the empty expression vector and included in every experiment. Artifacts due to bacterial lysis were excluded by using antisera raised against six different known cytoplasmic proteins, all of which gave negative results.

Regions of Atypical Nucleotide Composition.

These regions were identified by the χ^2 analysis: the distribution of all 64 trinucleotides (3 mers) was computed for the complete genome in all six reading frames, followed by the 3-mer distribution in 2,000-bp windows. Windows overlapped by 1,000 bp. For each window, the χ^2 statistic on the difference between its 3-mer content, and that of the whole genome was computed.

In Silico Genome Comparisons

The protein sets of *S. agalactiae*, *Streptococcus pneumoniae* and *S. pyogenes* were compared by using FASTA3. A general description of the FASTA3 sequence comparison program is discussed in Pearson, W.R., "Flexible Sequence Similarity Searching with the FASTA3 Program Package", (2000) *Methods Mol. Biol.*, 132: 185-219. Shared genes were defined using a FASTA3 *P* value cutoff of 10^{-15} . These shared genes and genes that *S. agalactiae* did not share with the other streptococci using this cutoff were subsequently searched against all completely sequenced genomes, and genes were defined as unique to streptococci or *S. agalactiae* when they did not share similarity with any other gene sets with a FASTA3 *P* value of 10^{-5} or lower. The use of two cutoffs provides for a more stringent analysis of shared or unique genes.

Figure 2 is a schematic representation of in silico comparisons between streptococci. The protein sets of GBS, *S. pn.*, and GAS were compared by using FASTA3. Numbers under the species name indicate genes that are not shared with the other species; values in parenthesis are the number of proteins in each species (excluding frame-shifted and degenerated genes). Numbers in the intersections indicate genes shared by two or three species. These are displayed in the color corresponding to the species used as the query. (GBS: green; *S.pn.*: blue; GAS: red). Numbers in any given intersection are slightly different due to gene duplications in some species.

Table 3 lists genes which were shared among GBS, GAS and pneumococcus, but which were not found in any of the other completely sequenced genomes. The protein sets of *S. agalactiae*, *S. pneumoniae*, and *S. pyogenes* were compared using FASTA3 [Pearson, W. R. (2000) *Methods Mol Biol* 132, 185-219]. Shared genes were defined using a FASTA3 p value cutoff of 10^{-15} . These shared genes and genes that *S. agalactiae* did not share with the other streptococci using this cutoff were subsequently searched against all completely sequenced genomes and genes were defined as unique to streptococci or *S. agalactiae* when they did not share similarity with any other gene sets with a FASTA3 p value of 10^{-5} or lower.

Synteny

Regions of conservation of gene synteny were computed as windows of 10 kb spanning at least three genes whose order was conserved in the other species. Regions were merged if they were less than 20 kb apart. The number of genes within each broad region was then calculated.

Comparative Genome Hybridizations

Comparative genome hybridizations (See Figure 1) using DNA microarrays were performed between the sequenced type V strain 2603 V/R and 19 other GBS strains of multiple serotypes (See Table %). Predicted genes from strain 2603 V/R

PATENT APPLICATION
ATTY REF NO. 19195.002

were amplified by PCR and arrayed on glass microscope slides. See Peterson, et al., (2000) *J. Bacteriol.* **182**, 6192-6202. Genomic DNA was labelled according to protocols provided by J. DeRisi (www.microarrays.org/Pdfs/Genomic-DNALabel_B.pdf), except that the DNA was not digested or sheared before labelling. Arrays were scanned with a GENEPIX 4000B scanner (Axon Instruments, Foster City, CA), and individual hybridisation signals were quantitated with TIGR SPOTFINDER. See Hedge, et al., (2000), *Biotechniques* **29**, 548-550, 552-554, 556. Cy3/Cy5 (2603 V/R signal/test strain) ratio cutoffs were defined arbitrarily as Cy3/Cy5 = 1.0 – 3.0, gene present in test strain; 3.0 – 10.0, ambiguous result; >10.0, gene absent. For ambiguous results, the gene may be divergent in the test strain relative to 2603 V/R, or the gene may be absent in the test strain but still produces paralogous gene family or a repetitive elemtn. Although cutoffs are arbitrary, they fit nicely the results for the variation of the capsule locus in the strains tested (see region 9 on Figure 1) where most genes are slightly divergent and only a few are completely different.

The CGH detected 1,698 genes in all of the strains, whereas 401 genes from strain 2603 V/R (18% of the gene complement) were not detected in at least one other strain, suggesting that they are absent or significantly divergent in those strains. Two hundred sixty (38%) of the 683 genes specific to *S. agalactiae* when compared with the other two streptococci (Fig. 2), including virulence determinants and surface proteins, vary among *S. agalactiae* strains, whereas only 47 (4%) of the genes common to all three streptococcal species, including 5 of the 6 sortases identified in the genome, vary among strains. Thus, the *in silico* analysis of genes shared by the streptococci that are not expected to vary among this genus is consistent with the CGH analysis. Forty-four (25%) of the genes shared by *S. agalactiae* and *S. pneumoniae* and 44 (20%) of those shared by *S. agalactiae* and *S. pyogenes* vary in the CGH analysis. The first set contains many glycosyl transferases and proteins carrying a cell-wall anchor, whereas the second set displays many phage-related genes. One hundred thirty-six of the 315 genes unique to *S. agalactiae* when

compared with all sequenced genomes vary among strains. These include R5, three capsular genes, two cell wall-anchored proteins, and three transcriptional regulators. Three hundred sixty-four (91%) of the 401 varying genes correspond to 15 regions containing more than 5 contiguous genes. Ten of these regions display an atypical nucleotide composition in strain 2603 V/R (Fig. 1), consistent with the possibility that they were horizontally transferred into this strain. Two of the largest regions (region 4, a prophage and region 7, similar to Tn916 from *Enterococcus faecalis*) are flanked by insertion sequence elements. The 15 regions contain many proteins predicted to be anchored on the cell wall or surface exposed, including Rib (region 3), sortases, glycosyl transferases, the capsule locus (region 9, divergent in all strains but the other type V strain CJB111), and phage-related genes. Region 14 is unique to *S. agalactiae* and spans 33 genes (SAG1989- SAG2021), including 25 proteins of unknown function, some of which carry a cell-wall anchor. It is flanked by an ISL3 transposase and displays an atypical nucleotide composition. Region 1, unique to *S. agalactiae*, is a possible plasmid or remnant of a phage (SAG0218-SAG0238), contains mostly hypothetical proteins, and is flanked by a site-specific recombinase. Region 8 is specific to *S. agalactiae*, comprises 20 proteins of unknown function (SAG1018-SAG1037), most of which are predicted to be membrane associated or secreted, and displays an atypical nucleotide composition.

The CGH results were analyzed by profile clustering where genes are grouped based on their distribution patterns (Fig. 5). Sixteen clusters of five or more contiguous and noncontiguous genes comprising a total of 300 genes were identified (Table 6). Several clusters correspond to regions of contiguous genes described above. Some clusters of genes that do not share sequence similarity and are located at different loci in the genome display an identical profile. For instance, a cluster of genes containing a surface antigen (SAG0674-SAG0681) follows the same distribution as another cluster containing only hypothetical proteins (SAG0247-

SAG0249). A putative pathogenicity protein (SAG2063) also clusters with a region containing several glycosyl transferases and Sec proteins (SAG1447-SAG1462).

Profile clustering was also used to group strains based on similarity of gene content (Fig. 5). In addition, the sequences of 19 genes from each of 11 *S. agalactiae* strains were determined after PCR amplification and used for phylogenetic analyses. The strains were the following: type Ia, 090 and A909; type Ib, H36B; type II, 18RS21; type III, COH1, M732 and M781; type V, 2603 V/R and 1169NT1; type VIII, JM9130013; and nontypeable strain CJB110. The set comprised 8 housekeeping genes and 11 genes coding for proteins predicted to be surface-exposed (Table 7).

The profile clustering was conducted as follows. The information and absence of genes based on the comparative genome hybridisation results was used to group genes based on their distribution patterns. The analysis used was essentially identical to that used for phylogenetic profile analysis. See Pellegrinie, et al., (1999) *Proc. Natl. Acad. Sci. USA* 96, 4285 – 4288. Each gene was assigned a binary profile based on its presence or absence across the different strains, with presence determined by a Cy3/Cy5 ratio < 3.0 and absence \geq 3.0. The gene profiles were then clustered by using the single-linkage clustering algorithm with column weighting (all with default settings) of CLUSTER (<http://rana.lbl.gov>). The CLUSTER program also groups the strains (columns) based on similarity of gene profiles. Clusters of genes and strains were viewed by using TREEVIEW (<http://rana.lbl.gov>).

Phylogenetic trees were inferred for the complete set of 19 genes and for the subsets of housekeeping and surface-exposed genes. Because the branching patterns in all three trees were identical, only the tree of the 19 genes is shown in Fig. 3. The degree of polymorphism of the housekeeping and the surface-exposed genes is similar (~1 variable site among all of the strains per 100 bp).

The sequences of genes from the different strains were aligned by using CLUSTALW (See Thompson (1994), *Nucleic Acids Res.* 22, 4673 – 4680.) and

PATENT APPLICATION
ATTY REF NO. 19195.002

trimmed to remove ambiguously aligned regions. Phylogenetic trees of individual genes and of concatenated alignments of multiple genes were inferred by using maximum likelihood methods of PAUP* 4.0 b10 (Sinauer, Sunderland, MA). Bootstrap analysis was carried out using PAUP* as well. The possibility of recombination among strains was examined by using analysis of sequence variation using SIMPLOT (S.C. Ray) and analysis of phylogenetic heterogeneity by using MACCLADE (Sinauer).

Analysis of this variation showed no evidence for major recombination events between the strains. There were no long stretches of polymorphic sites that strongly supported other trees (analysis with MACCLADE), and there were no significant crossover events in plots of sequence similarity between strains (analysis with SIMPLOT). Some strain groupings (clades) generated by phylogenetic analysis were similar to clusters from the profile analysis (type III strains M781, M732 and COH1; type Ia strain 090 and nontypable strain CJB110), whereas others were different, possibly because of the aforementioned problems with the profile clustering. In both the phylogenetic analysis and the profile clustering, there is serotypedependent and -independent clustering (Figs. 3 and 5). The presence of strains of the same serotype in different clades or clusters could be due to lateral gene transfer.

Figure 5 demonstrates phylogenetic profiling of GBS strains based on comparative genome hybridisations. The information on presence and absence of genes based on the microarray comparative genome hybridization results was used for phylogenetic profile analysis. The presence of a particular gene or gene cluster is indicated in the figure by a red square and the absence of a gene or cluster by a black square. The relationship between strains based on this analysis is depicted by the tree at the top of the figure. The strains and their serotypes are indicated (NT: nontypeable). Clusters with identical profiles are reduced to a single horizontal line and the number of genes in each cluster is indicated on the right. The clusters of 5 or more genes, labeled in red text and numbered, are listed in Table 6. The 1698 genes shared by all 19 strains are labeled in green text.

Figure 3 depicts a phylogenetic tree of GBS strains based on PCR sequences. The sequences of 19 genes (Table 7) from each of 11 GBS strains were aligned and trimmed to remove ambiguously aligned regions, and phylogenetic trees were inferred. Strain names are indicated in bold, and serotypes are indicated under the strain names. Bootstrap values are indicated on the branches.

Techniques

A summary of standard techniques and procedures which may be employed in order to perform the invention (*e.g.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention, but gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989) or *Third Edition* (2000); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

Further Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "an epithelial cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, *etc.*

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Streptococcal sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the

mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

Expression systems

The Streptococcal nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells. The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence

fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicon systems, thus allowing it to be maintained, for example, in mammalian cells for

expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San

Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers & Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) ("Summers & Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extra-chromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

PATENT APPLICATION
ATTY REF NO. 19195.002

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers & Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers & Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, etc. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also present in the medium, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991).

Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987).

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for

example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Rept.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985.

The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*,

Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called

an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g-laotamase* (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled

PATENT APPLICATION
ATTY REF NO. 19195.002

with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be

PATENT APPLICATION
ATTY REF NO. 19195.002

made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698]. Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in

terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes. Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A-0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See eg. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem.* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ.*

Microbiol. 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, Streptococcus].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GALA*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter

can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved

either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YE_p24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number

ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg. Brake et al., supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al., supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

PATENT APPLICATION
ATTY REF NO. 19195.002

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcal proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired,

the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of

PATENT APPLICATION
ATTY REF NO. 19195.002

pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated. Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule. See also *Delivery Strategies for Antisense Oligonucleotide Therapeutics* (ed. Akhtar) ISBN 0849347785.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized

PATENT APPLICATION
ATTY REF NO. 19195.002

macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Additionally, these carriers may function as immunostimulating agents ("adjuvants").

Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to:

(1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be used [eg. Robinson & Torres (1997) *Seminars in Immunol* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

PATENT APPLICATION
ATTY REF NO. 19195.002

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum. Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res*

PATENT APPLICATION
ATTY REF NO. 19195.002

33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further

PATENT APPLICATION
ATTY REF NO. 19195.002

example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC with accession numbers VR-977 and VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

PATENT APPLICATION
ATTY REF NO. 19195.002

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

PATENT APPLICATION
ATTY REF NO. 19195.002

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin. Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be

incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

PATENT APPLICATION
ATTY REF NO. 19195.002

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hypodermic syringes. Dosage treatment may be a single dose schedule or a multiple dose schedule. Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

The terms "polynucleotide" and "nucleic acid", used interchangeably herein, In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

One example are polypeptides which include, without limitation: asialoglycoprotein (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

PATENT APPLICATION
ATTY REF NO. 19195.002

B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527. Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example,

N[1-(2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and

DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C & E, over time these lipoproteins lose A and acquire C & E. VLDL comprises A, B, C & E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, *etc.*

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin, and lipofectAMINE are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Streptococcus antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Streptococcus antibodies can be used to detect antigen levels).

Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Streptococcus proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is

subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Use of Polypeptides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full length genes can be used to screen peptide libraries to identify binding partners, such as receptors, from within the library. Peptide libraries can be synthesized according to methods known in the art (*e.g.* U.S. patent 5,010,175; WO91/17823). Agonists or antagonists of the polypeptides of the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, *etc.* The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide described herein, and at least one peptide agonist or antagonist of the binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

Identification of anti-bacterial agents

Drug Screening Assays

Of particular interest in the present invention is the identification of agents that have activity in modulating expression of one or more of the adhesion-specific genes described herein, so as to inhibit infection and/or disease. Of particular interest are screening assays for agents that have a low toxicity for human cells.

The term "agent" as used herein describes any molecule with the capability of altering or mimicking the expression or physiological function of a gene product of a differentially expressed gene. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, including, but not limited to, organic molecules (*e.g.* small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons), peptides, antisense polynucleotides, and ribozymes, and the like.

Candidate agents can comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate

agents are also found among biomolecules including, but not limited to: polynucleotides, peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Screening of Candidate Agents In Vitro

A wide variety of *in vitro* assays may be used to screen candidate agents for the desired biological activity, including, but not limited to, labeled *in vitro* protein-protein binding assays, protein-DNA binding assays (*e.g.* to identify agents that affect expression), electrophoretic mobility shift assays, immunoassays for protein binding, and the like. For example, by providing for the production of large amounts of a differentially expressed polypeptide, one can identify ligands or substrates that bind to, modulate or mimic the action of the polypeptide. The purified polypeptide may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, transcriptional regulation, *etc.* The screening assay can be a binding assay, wherein one or more of the molecules may be joined to a label, and the label directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemilumescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin *etc.* For the specific binding members, the

PATENT APPLICATION
ATTY REF NO. 19195.002

complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assays described herein. Where the assay is a binding assay, these include reagents like salts, neutral proteins, *e.g.* albumin, detergents, *etc.* that are used to facilitate optimal protein-protein binding, protein-DNA binding, and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40°C.

Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

Many mammalian genes have homologs in yeast and lower animals. The study of such homologs' physiological role and interactions with other proteins *in vivo* or *in vitro* can facilitate understanding of biological function. In addition to model systems based on genetic complementation, yeast has been shown to be a powerful tool for studying protein-protein interactions through the two hybrid system.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

PATENT APPLICATION
ATTY REF NO. 19195.002

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the *Streptococcus* nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native *Streptococcal* sequence is preferred

because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Streptococcal sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Streptococcus sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Streptococcus sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Streptococcus sequence in order to hybridize therewith and thereby form a duplex which can be detected. The exact length and sequence of the probe will depend on the hybridization conditions (*e.g.* temperature, salt condition *etc.*). For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH*

14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acid. The assay is described in Mullis *et al.* [*Meth. Enzymol.* (1987) 155:335-350] & US patents 4,683,195 & 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site.

Typically, such sequence will flank the desired *Streptococcus* sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the *Streptococcus* sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

PATENT APPLICATION
ATTY REF NO. 19195.002

ABSTRACT

The invention relates to polynucleotides which are conserved or specific to one or more species of *Streptococcus*, *Streptococcus* species serotypes, and/or serotype isolates. In particular, the invention relates to polynucleotides from *Streptococcus* which are conserved or specific to one or more of the species of *S. pneumoniae* ("pneumococcus" or "S. pn."), *S. pyogenes* ("group A streptococcus" or "GAS"), and *S. agalactiae* ("group B streptococcus" or "GBS"). The invention further relates to polynucleotides which are conserved or specific to one or more *Streptococcal* species serotypes, such as GBS serotypes Ia, Ib, II, III, IV, V, VI, VII, and VIII. The invention still further relates to polynucleotides which are conserved or specific to one or more clinical isolates of a *Streptococcus* species.

Figure 1

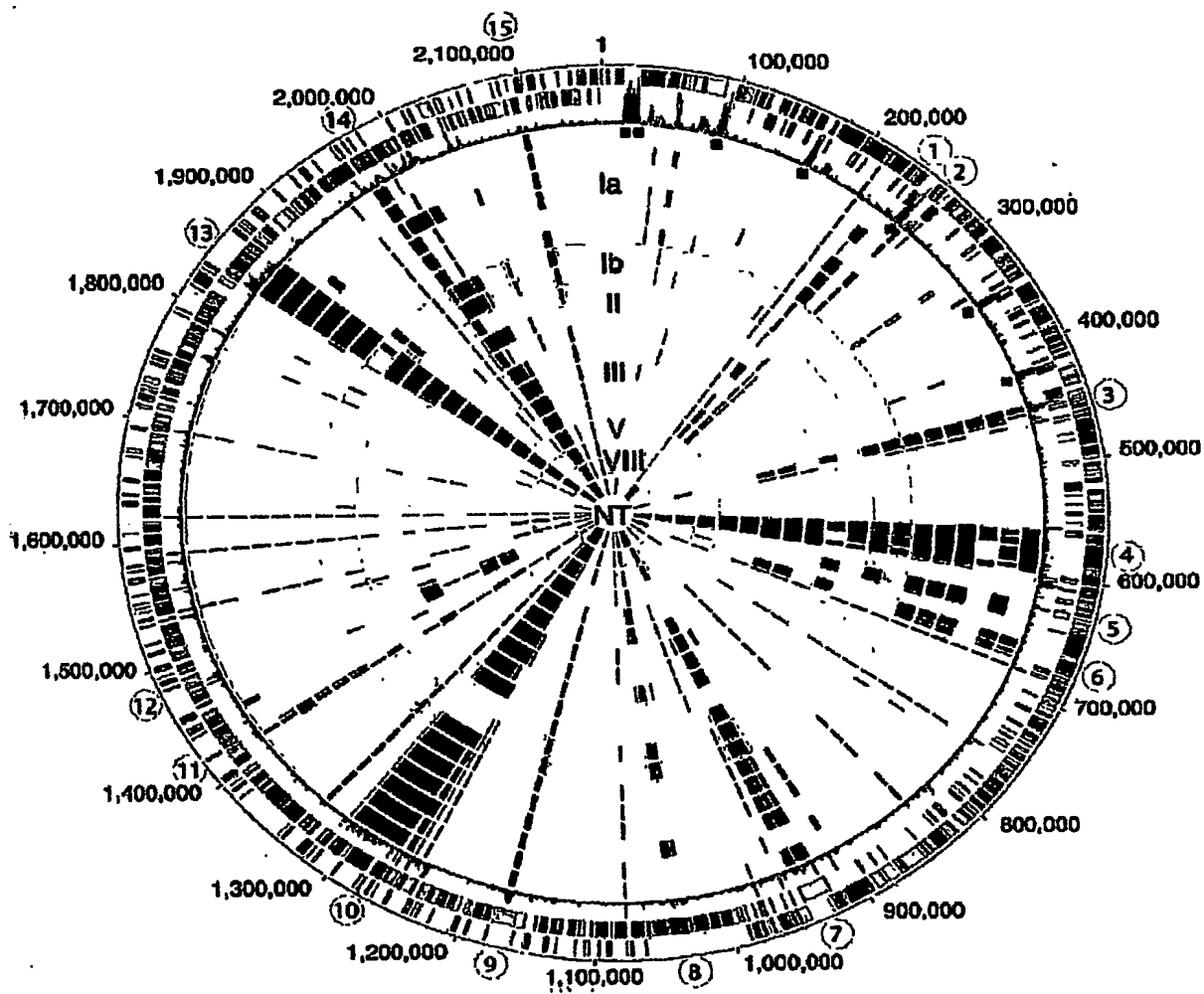


Figure 2

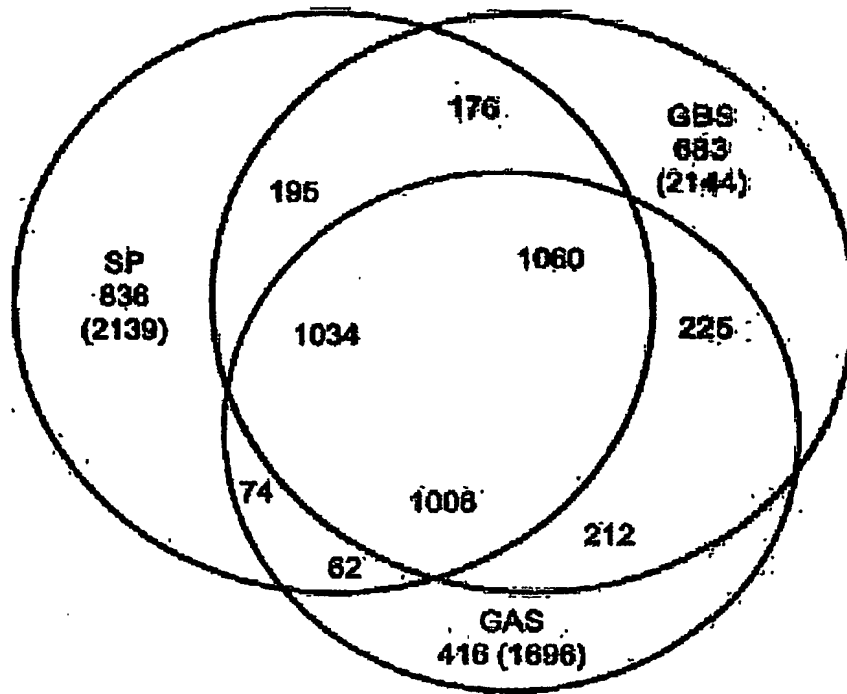


Figure 3

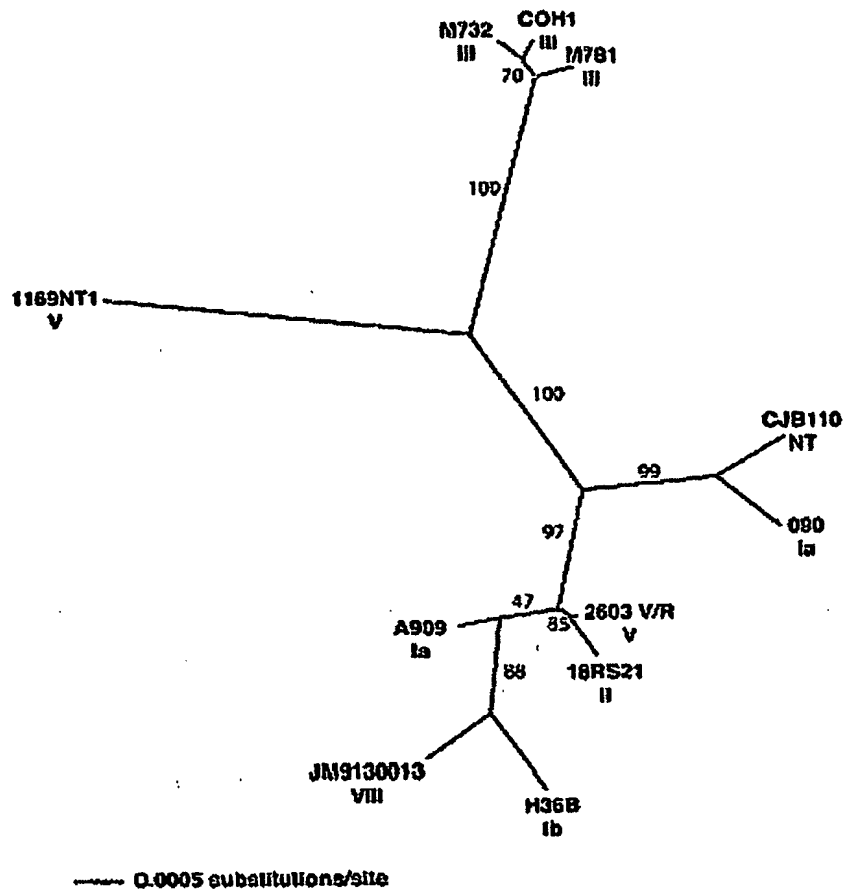


Figure 5

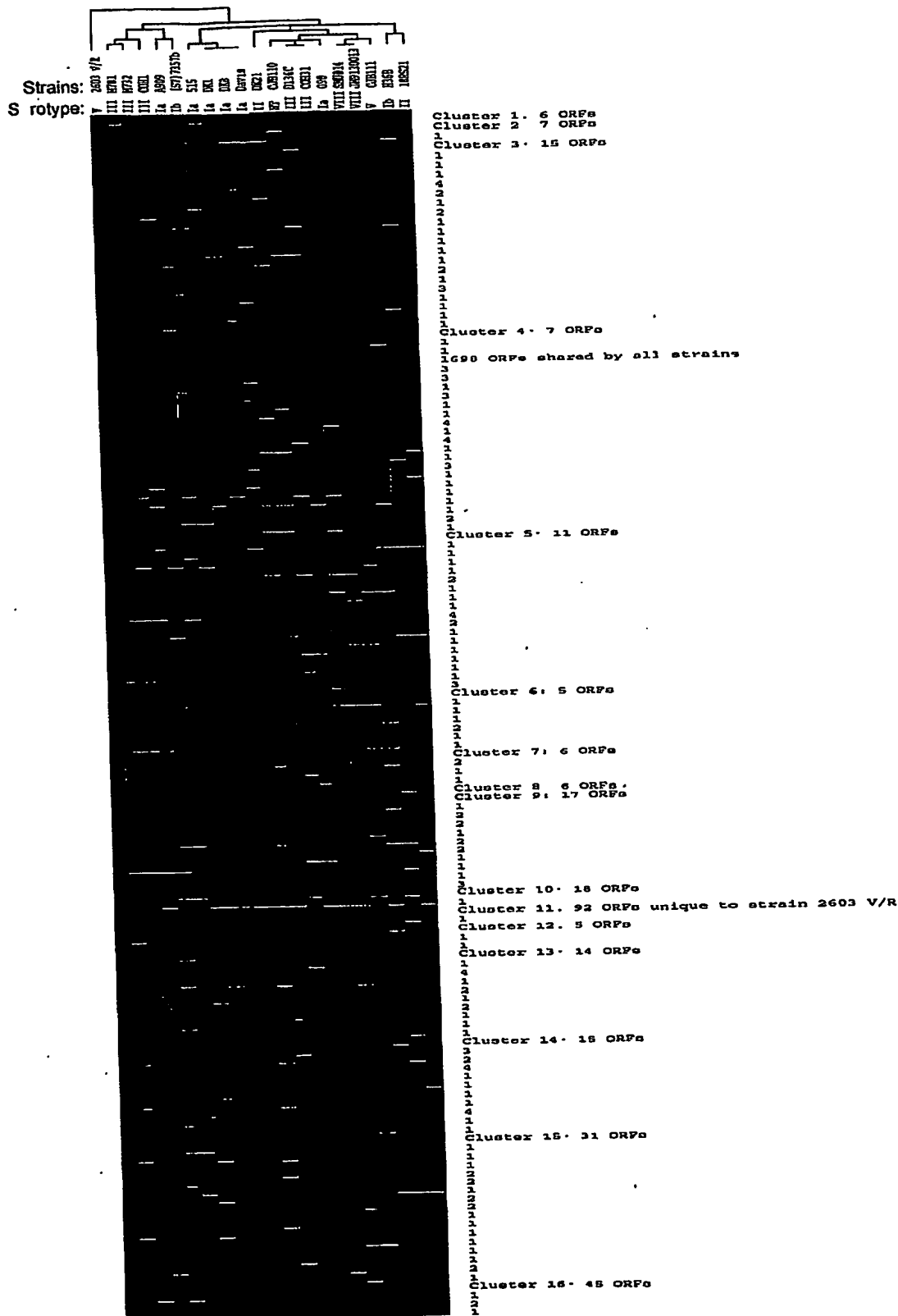


Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0001	453	chromosomal replication initiator protein DnaA
SAG0002	378	DNA polymerase III, beta subunit
SAG0003	293	diacylglycerol kinase catalytic domain protein, putative
SAG0004	65	conserved hypothetical protein
SAG0005	67	hypothetical protein
SAG0006	371	GTP-binding protein YchF
SAG0007	191	peptidyl-tRNA hydrolase
SAG0008	1165	transcription-repair coupling factor
SAG0009	31	hypothetical protein
SAG0010	90	S4 domain protein
SAG0011	123	cell division protein DivIC, putative
SAG0012	44	conserved hypothetical protein
SAG0013	428	protein of unknown function
SAG0014	424	MesJ/Ycf62 family protein
SAG0015	180	hypoxanthine-guanine phosphoribosyltransferase
SAG0016	658	cell division protein FtsH
SAG0017	447	pcsB protein
SAG0018	322	ribose-phosphate pyrophosphokinase
SAG0019	391	aminotransferase, class I
SAG0020	253	recombination protein O
SAG0021	283	protease, putative
SAG0022	330	fatty acid/phospholipid synthesis protein PlsX
SAG0023	79	acyl carrier protein
SAG0024	234	phosphoribosylaminoimidazole-succinocarboxamide synthase
SAG0025	1241	phosphoribosylformylglycinamide synthase, putative
SAG0026	484	amidophosphoribosyltransferase
SAG0027	340	phosphoribosylformylglycinamide cyclo-ligase
SAG0028	182	phosphoribosylglycinamide formyltransferase
SAG0029	250	acetyltransferase, GNAT family
SAG0030	515	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase
SAG0031	299	peptidase, M23/M37 family
SAG0032	434	group B streptococcal surface immunogenic protein
SAG0033	232	N-acetylmannosamine-6-P epimerase, putative
SAG0034	438	sugar ABC transporter, sugar-binding protein
SAG0035	295	sugar ABC transporter, permease protein
SAG0036	276	sugar ABC transporter, permease protein
SAG0037	147	conserved hypothetical protein
SAG0038	220	conserved hypothetical protein
SAG0039	305	N-acetylneuraminate lyase, putative
SAG0040	293	ROK family protein
SAG0041	325	acetyl xylan esterase, putative
SAG0042	267	phosphosugar-binding transcriptional regulator, RpiR family, putative
SAG0043	421	phosphoribosylamine--glycine ligase
SAG0044	162	phosphoribosylaminoimidazole carboxylase, catalytic subunit
SAG0045	363	phosphoribosylaminoimidazole carboxylase, ATPase subunit
SAG0046	463	membrane protein, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0047	432	adenylosuccinate lyase
SAG0048	303	transcriptional regulator, Cro/CI family
SAG0049	332	Holliday junction DNA helicase RuvB
SAG0050	145	phosphotyrosine protein phosphatase, low molecular weight
SAG0051	126	MORN motif family protein
SAG0052	592	membrane protein, putative
SAG0053	880	aldehyde-alcohol dehydrogenase
SAG0054	338	alcohol dehydrogenase, propanol-preferring
SAG0055	496	threonine synthase
SAG0056	412	MATE efflux family protein
SAG0057	102	ribosomal protein S10
SAG0058	208	ribosomal protein L3
SAG0059	207	ribosomal protein L4
SAG0060	98	ribosomal protein L23
SAG0061	277	ribosomal protein L2
SAG0062	92	ribosomal protein S19
SAG0063	114	ribosomal protein L22
SAG0064	217	ribosomal protein S3
SAG0065	137	ribosomal protein L16
SAG0066	68	ribosomal protein L29
SAG0067	86	ribosomal protein S17
SAG0068	122	ribosomal protein L14
SAG0069	101	ribosomal protein L24
SAG0070	180	ribosomal protein L5
SAG0071	61	ribosomal protein S14, putative
SAG0072	132	ribosomal protein S8
SAG0073	178	ribosomal protein L6
SAG0074	118	ribosomal protein L18
SAG0075	164	ribosomal protein S5
SAG0076	59	ribosomal protein L30
SAG0077	146	ribosomal protein L15
SAG0078	434	preprotein translocase, SecY subunit
SAG0079	212	adenylate kinase
SAG0080	72	translation initiation factor IF-1
SAG0081	38	ribosomal protein L36
SAG0082	121	ribosomal protein S13
SAG0083	118	ribosomal protein S11
SAG0084	312	DNA-directed RNA polymerase, alpha subunit
SAG0085	128	ribosomal protein L17
SAG0086	85	lipoprotein, putative
SAG0087	59	hypothetical protein
SAG0088	56	hypothetical protein
SAG0089	183	conserved hypothetical protein
SAG0090	139	conserved hypothetical protein
SAG0091	144	transcriptional regulator ComX1, putative
SAG0092	230	phosphoglycerate mutase family protein
SAG0093	250	D-alanyl-D-alanine carboxypeptidase family protein
SAG0094	191	N-acetylmuramoyl-L-alanine amidase, family 4 protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0095	344	heat-inducible transcription repressor HrcA
SAG0096	190	heat shock protein GrpE
SAG0097	609	dnaK protein
SAG0098	379	dnaJ protein
SAG0099	415	transcriptional regulator, GntR family
SAG0100	258	tRNA pseudouridine synthase A
SAG0101	252	phosphomethylpyrimidine kinase, putative
SAG0102	154	conserved hypothetical protein
SAG0103	189	conserved hypothetical protein TIGR01440
SAG0104	280	conserved hypothetical protein
SAG0105	427	trigger factor
SAG0106	191	DNA-directed RNA polymerase, delta subunit, putative
SAG0107	534	CTP synthase
SAG0108	308	conserved hypothetical protein
SAG0109	148	deoxyuridine 5'-triphosphate nucleotidohydrolase
SAG0110	454	DNA repair protein RadA
SAG0111	165	carbonic anhydrase-related protein
SAG0112	439	pyridine nucleotide-disulphide oxidoreductase family protein
SAG0113	484	glutamyl-tRNA synthetase
SAG0114	322	ribose ABC transporter, periplasmic D-ribose-binding protein
SAG0115	310	ribose ABC transporter, permease protein
SAG0116	492	ribose ABC transporter, ATP-binding protein
SAG0117	132	ribose ABC transporter protein RbsD
SAG0118	303	ribokinase
SAG0119	328	ribose operon repressor RbsR
SAG0120	32	hypothetical protein
SAG0121	362	permease, putative
SAG0122	228	ABC transporter, ATP-binding protein
SAG0123	223	DNA-binding response regulator
SAG0124	356	sensor histidine kinase
SAG0125	396	argininosuccinate synthase
SAG0126	462	argininosuccinate lyase
SAG0127	293	fructose-bisphosphate aldolase
SAG0128	305	L-2-hydroxyisocaproate dehydrogenase
SAG0129	62	ribosomal protein L28
SAG0130	121	conserved hypothetical protein
SAG0131	543	DAK2 domain protein
SAG0132	294	SPFH domain/Band 7 family protein
SAG0133	38	conserved hypothetical protein
SAG0134	96	hypothetical protein
SAG0135	246	amino acid ABC transporter, ATP-binding protein
SAG0136	516	amino acid ABC transporter, amino acid-binding protein/permease protein
SAG0137	627	conserved hypothetical protein
SAG0138	279	undecaprenol kinase, putative
SAG0139	251	negative regulator of competence MecA, putative
SAG0140	386	glycosyl transferase, group 4 family protein
SAG0141	256	ABC transporter, ATP-binding protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0142	420	conserved hypothetical protein
SAG0143	410	selenocysteine lyase
SAG0144	147	NifU family protein
SAG0145	472	conserved hypothetical protein
SAG0146	395	penicillin-binding protein 4, putative
SAG0147	411	D-alanyl-D-alanine carboxypeptidase family protein
SAG0148	551	oligopeptide ABC transporter, substrate-binding protein, putative
SAG0149	304	oligopeptide ABC transporter, permease protein
SAG0150	343	oligopeptide ABC transporter, permease protein
SAG0151	348	oligopeptide ABC transporter, ATP-binding protein
SAG0152	310	oligopeptide ABC transporter, ATP-binding protein
SAG0153	283	4-diphosphocytidyl-2C-methyl-D-erythritol kinase
SAG0154	147	adc operon repressor AdcR
SAG0155	236	zinc ABC transporter, ATP-binding protein
SAG0156	270	zinc ABC transporter, permease protein
SAG0157	NA	deoxyribonuclease-related protein, degenerate
SAG0158	419	tyrosyl-tRNA synthetase
SAG0159	765	penicillin-binding protein 1B, putative
SAG0160	1191	DNA-directed RNA polymerase, beta subunit
SAG0161	1216	DNA-directed RNA polymerase beta' subunit
SAG0162	121	conserved hypothetical protein
SAG0163	323	competence protein CglA
SAG0164	282	competence protein CglB
SAG0165	151	conserved hypothetical protein
SAG0166	123	conserved domain protein
SAG0167	324	conserved hypothetical protein
SAG0168	397	acetate kinase
SAG0169	68	transcriptional regulator, Cro/CI family
SAG0170	45	hypothetical protein
SAG0171	151	hypothetical protein
SAG0172	221	protease, putative
SAG0173	256	pyrroline-5-carboxylate reductase
SAG0174	355	glutamyl-aminopeptidase
SAG0175	79	hypothetical protein
SAG0176	94	conserved hypothetical protein
SAG0177	107	thioredoxin family protein
SAG0178	208	tRNA binding domain protein
SAG0179	238	conserved hypothetical protein
SAG0180	131	single-strand binding protein
SAG0181	214	hydrolase, haloacid dehalogenase-like family
SAG0182	581	sensor histidine kinase, putative
SAG0183	246	response regulator
SAG0184	151	conserved hypothetical protein
SAG0185	242	membrane protein, putative
SAG0186	36	hypothetical protein
SAG0187	542	oligopeptide ABC transporter, oligopeptide-binding protein
SAG0188	325	oligopeptide ABC transporter, permease protein
SAG0189	273	oligopeptide ABC transporter, permease protein

Tabl 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0190	267	peptide ABC transporter, ATP-binding protein
SAG0191	208	peptide ABC transporter, ATP-binding protein
SAG0192	676	PTS system, IIABC components
SAG0193	541	alpha amylase family protein
SAG0194	639	transcriptional antiterminator, BglG family
SAG0195	377	IS1548, transposase
SAG0196	66	conserved domain protein
SAG0197	94	PTS system, IIB component, putative
SAG0198	451	PTS system, IIC component, putative
SAG0199	285	transketolase, N-terminal subunit
SAG0200	309	transketolase, C-terminal subunit
SAG0201	419	oxidoreductase, putative
SAG0202	89	ribosomal protein S15
SAG0203	709	polyribonucleotide nucleotidyltransferase
SAG0204	250	conserved hypothetical protein
SAG0205	194	serine O-acetyltransferase
SAG0206	60	lipoprotein, putative
SAG0207	447	cysteinyI-tRNA synthetase
SAG0208	128	conserved hypothetical protein
SAG0209	251	RNA methyltransferase, TrmH family, group 3
SAG0210	172	conserved hypothetical protein
SAG0211	286	DegV family protein
SAG0212	32	hypothetical protein
SAG0213	39	hypothetical protein
SAG0214	148	ribosomal protein L13
SAG0215	130	ribosomal protein S9
SAG0216	33	hypothetical protein
SAG0217	384	site-specific recombinase, phage integrase family
SAG0218	158	transcriptional regulator, Cro/CI family
SAG0219	101	hypothetical protein
SAG0220	92	conserved hypothetical protein
SAG0221	76	hypothetical protein
SAG0222	108	conserved domain protein
SAG0223	209	conserved hypothetical protein, fusion
SAG0224	332	replication initiation protein, putative
SAG0225	144	hypothetical protein
SAG0226	418	recombination protein
SAG0227	156	hypothetical protein
SAG0228	111	conserved hypothetical protein
SAG0229	95	conserved hypothetical protein
SAG0230	96	conserved hypothetical protein
SAG0231	135	hypothetical protein
SAG0232	186	hypothetical protein
SAG0233	226	hypothetical protein
SAG0234	128	hypothetical protein
SAG0235	93	hypothetical protein
SAG0236	32	hypothetical protein
SAG0237	34	hypothetical protein

Table 1: Complet list of GBS predicted genes

ORF	Siz (a.a.)	Annotation
SAG0238	41	hypothetical protein
SAG0239	286	transcriptional regulator MutR family
SAG0240	393	transporter, putative
SAG0241	213	amino acid ABC transporter, permease protein
SAG0242	308	amino acid ABC transporter, amino acid-binding protein
SAG0243	211	amino acid ABC transporter, permease protein
SAG0244	381	amino acid ABC transporter, ATP-binding protein
SAG0245	152	protein of unknown function/lipoprotein, putative
SAG0246	268	hypothetical protein
SAG0247	116	hypothetical protein
SAG0248	90	hypothetical protein
SAG0249	116	hypothetical protein
SAG0250	193	membrane protein, putative
SAG0251	72	transcriptional regulator, Cro/CI family
SAG0252	186	acetyltransferase, GNAT family
SAG0253	192	acetyltransferase, GNAT family
SAG0254	226	acetyltransferase, GNAT family
SAG0255	315	conserved hypothetical protein
SAG0256	163	RNA polymerase sigma factor, ECF subfamily
SAG0257	53	lipoprotein, putative
SAG0258	202	transcriptional regulator, TetR family
SAG0259	365	ABC transporter efflux protein, DrrB family, putative
SAG0260	238	ABC transporter, ATP-binding protein
SAG0261	129	IS1381, transposase OrfB
SAG0262	127	IS1381, transposase OrfA
SAG0263	171	hypothetical protein
SAG0264	103	conserved hypothetical protein
SAG0265	235	conserved hypothetical protein
SAG0266	382	N-acetylglucosamine-6-phosphate deacetylase
SAG0267	180	conserved hypothetical protein
SAG0268	304	glycyl-tRNA synthetase, alpha subunit
SAG0269	213	acyl carrier protein phosphodiesterase, putative
SAG0270	679	glycyl-tRNA synthetase, beta subunit
SAG0271	85	conserved hypothetical protein
SAG0272	87	membrane protein, putative
SAG0273	502	glycerol kinase
SAG0274	609	alpha-glycerophosphate oxidase
SAG0275	232	glycerol uptake facilitator protein
SAG0276	445	NADH oxidase, putative
SAG0277	476	conserved hypothetical protein
SAG0278	661	transketolase
SAG0279	101	conserved hypothetical protein
SAG0280	244	ABC transporter, ATP-binding protein
SAG0281	534	membrane protein, putative
SAG0282	461	PTS system, IIBC components
SAG0283	267	glutamate 5-kinase
SAG0284	417	gamma-glutamyl phosphate reductase
SAG0285	298	conserved hypothetical protein TIGR00006

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0286	108	cell division protein FtsL, putative
SAG0287	752	penicillin-binding protein 2X
SAG0288	336	phospho-N-acetylmuramoyl-pentapeptide-transferase
SAG0289	447	ATP-dependent RNA helicase, DEAD/DEAH box family
SAG0290	270	ABC transporter, substrate-binding protein
SAG0291	267	amino acid ABC transporter, permease protein
SAG0292	247	amino acid ABC transporter, ATP-binding protein
SAG0293	74	conserved hypothetical protein
SAG0294	304	thioredoxin reductase
SAG0295	486	conserved hypothetical protein
SAG0296	273	NAD synthetase
SAG0297	444	aminopeptidase C
SAG0298	750	penicillin-binding protein 1A
SAG0299	199	recombination protein U
SAG0300	172	conserved hypothetical protein
SAG0301	40	hypothetical protein
SAG0302	110	conserved hypothetical protein
SAG0303	384	conserved hypothetical protein
SAG0304	487	conserved hypothetical protein
SAG0305	160	autoinducer-2 production protein LuxS
SAG0306	535	KH domain protein
SAG0307	33	hypothetical protein
SAG0308	298	ABC transporter, ATP-binding protein
SAG0309	246	ABC transporter, permease protein, putative
SAG0310	361	conserved hypothetical protein
SAG0311	NA	DNA-binding response regulator, authentic point mutation
SAG0312	234	conserved hypothetical protein
SAG0313	209	guanylate kinase
SAG0314	104	DNA-directed RNA polymerase, omega subunit, putative
SAG0315	796	primosomal protein N'
SAG0316	311	methionyl-tRNA formyltransferase
SAG0317	440	sun protein
SAG0318	245	serine/threonine phosphatase, putative
SAG0319	651	serine/threonine protein kinase
SAG0320	231	conserved hypothetical protein
SAG0321	339	sensor histidine kinase, putative
SAG0322	213	DNA-binding response regulator
SAG0323	466	hydrolase, haloacid dehalogenase family/peptidyl-prolyl cis-trans isomerase, cyclophilin type
SAG0324	124	general stress protein, putative
SAG0325	258	pyruvate formate-lyase-activating enzyme
SAG0326	251	transcriptional regulator, DeoR family
SAG0327	327	transcriptional regulator, putative
SAG0328	107	PTS system, cellobiose-specific IIA component
SAG0329	106	PTS system, cellobiose-specific IIB component
SAG0330	433	PTS system, cellobiose-specific IIC component
SAG0331	818	formate acetyltransferase
SAG0332	222	transaldolase family protein

Table 1: Complet list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0333	362	glycerol dehydrogenase
SAG0334	308	cysteine synthase A
SAG0335	214	conserved hypothetical protein TIGR00257
SAG0336	429	helicase, putative
SAG0337	221	competence protein F, putative
SAG0338	184	ribosomal subunit interface protein
SAG0339	450	aspartate kinase family protein
SAG0340	216	hydrolase, haloacid dehalogenase-like family
SAG0341	49	hypothetical protein
SAG0342	263	enoyl-CoA hydratase/isomerase family protein
SAG0343	144	transcriptional regulator, MarR family
SAG0344	323	3-oxoacyl-(acyl-carrier-protein) synthase III
SAG0345	74	acyl carrier protein
SAG0346	319	enoyl-(acyl-carrier-protein) reductase II
SAG0347	308	malonyl CoA-acyl carrier protein transacylase
SAG0348	244	3-oxoacyl-[acyl-carrier protein] reductase
SAG0349	410	3-oxoacyl-(acyl-carrier-protein) synthase II
SAG0350	166	acetyl-CoA carboxylase, biotin carboxyl carrier protein
SAG0351	140	(3R)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase
SAG0352	456	acetyl-CoA carboxylase, biotin carboxylase
SAG0353	291	acetyl-CoA carboxylase, carboxyl transferase, beta subunit
SAG0354	257	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
SAG0355	210	conserved hypothetical protein
SAG0356	425	seryl-tRNA synthetase
SAG0357	330	membrane protein, putative
SAG0358	120	conserved hypothetical protein
SAG0359	303	PTS system, mannose-specific IID component
SAG0360	270	PTS system, mannose-specific IIC component
SAG0361	336	PTS system, mannose-specific IIAB components
SAG0362	270	hydrolase, haloacid dehalogenase-like family
SAG0363	194	hypothetical protein
SAG0364	203	membrane protein, putative
SAG0365	473	xanthine/uracil permease family protein
SAG0366	169	conserved hypothetical protein TIGR00150
SAG0367	186	acetyltransferase, GNAT family
SAG0368	435	protein of unknown function
SAG0369	98	conserved hypothetical protein
SAG0370	139	HIT family protein
SAG0371	167	hypothetical protein
SAG0372	85	hypothetical protein
SAG0373	241	ABC transporter, ATP-binding protein
SAG0374	344	ABC transporter, permease protein
SAG0375	266	conserved hypothetical protein
SAG0376	211	conserved hypothetical protein TIGR00091
SAG0377	127	conserved hypothetical protein
SAG0378	379	N utilization substance protein A
SAG0379	98	conserved hypothetical protein
SAG0380	100	ribosomal protein L7A family

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0381	927	translation initiation factor IF-2
SAG0382	122	ribosome-binding factor A
SAG0383	334	protein of unknown function/lipoprotein, putative
SAG0384	138	transcriptional repressor CopY
SAG0385	744	copper-transporter ATPase CopA
SAG0386	68	copper-transporter protein CopZ
SAG0387	204	membrane protein, putative
SAG0388	270	hydrolase, haloacid dehalogenase-like family
SAG0389	880	DNA polymerase I
SAG0390	146	CoA-binding domain protein
SAG0391	159	transcriptional regulator, Fur family
SAG0392	521	cell wall surface anchor family protein
SAG0393	228	DNA-binding response regulator
SAG0394	345	sensor histidine kinase
SAG0395	246	membrane protein, putative
SAG0396	380	queuine tRNA-ribosyltransferase
SAG0397	102	conserved hypothetical protein
SAG0398	179	BioY family protein
SAG0399	258	AtsA/ElaC family protein
SAG0400	168	cytidine/deoxycytidylate deaminase family protein
SAG0401	44	hypothetical protein
SAG0402	449	glucose-6-phosphate isomerase
SAG0403	175	5-formyltetrahydrofolate cyclo-ligase family protein
SAG0404	225	rhomboid family protein
SAG0405	347	protein of unknown function/lipoprotein, putative
SAG0406	299	UTP-glucose-1-phosphate uridylyltransferase
SAG0407	338	glycerol-3-phosphate dehydrogenase (NAD(P)+)
SAG0408	109	ribonuclease P protein component
SAG0409	271	SpoIIIJ family protein
SAG0410	273	R3H domain protein
SAG0411	177	conserved hypothetical protein
SAG0412	258	recX protein
SAG0413	451	RNA methyltransferase, TrmA family
SAG0414	153	conserved hypothetical protein
SAG0415	142	acetyltransferase, GNAT family
SAG0416	1233	protease, putative
SAG0417	302	glycosyl transferase, group 2 family protein
SAG0418	336	ribonucleoside-diphosphate reductase 2, beta subunit
SAG0419	137	nrdI protein
SAG0420	721	ribonucleoside-diphosphate reductase 2, alpha subunit
SAG0421	1055	cell wall surface anchor family protein
SAG0422	129	conserved hypothetical protein
SAG0423	132	conserved domain protein
SAG0424	94	hypothetical protein
SAG0425	105	carboxymuconolactone decarboxylase family protein
SAG0426	131	conserved hypothetical protein
SAG0427	129	transcriptional regulator, MerR family
SAG0428	345	alcohol dehydrogenase, zinc-containing

Table 1: C mplete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0429	284	oxidoreductase, aldo/keto reductase family
SAG0430	287	cation efflux system protein
SAG0431	174	transcriptional regulator, TetR family
SAG0432	397	transcriptional regulator, AraC family
SAG0433	1389	surface protein Rib
SAG0434	61	transposase, IS256 family, truncation
SAG0435	97	DNA-damage-inducible protein J, putative
SAG0436	62	hypothetical protein
SAG0437	123	lipoprotein, putative
SAG0438	145	bacteriophage L54a, integrase, truncation
SAG0439	NA	conserved hypothetical protein, degenerate
SAG0440	84	conserved hypothetical protein
SAG0441	103	conserved domain protein
SAG0442	189	acetyltransferase, GNAT family
SAG0443	194	acetyltransferase, GNAT family
SAG0444	188	conserved hypothetical protein
SAG0445	883	valyl-tRNA synthetase
SAG0446	319	oxidoreductase, Gfo/Idh/MocA family
SAG0447	287	magnesium transporter, CorA family
SAG0448	391	transposase, IS256 family
SAG0449	354	conserved hypothetical protein
SAG0450	330	aspartate--ammonia ligase
SAG0451	149	bacteriocin transport accessory protein, putative
SAG0452	179	type II DNA modification methyltransferase, putative
SAG0453	96	hypothetical protein
SAG0454	161	phosphopantetheine adenylyltransferase
SAG0455	357	conserved hypothetical protein
SAG0456	NA	conserved hypothetical protein, degenerate
SAG0457	192	conserved hypothetical protein
SAG0458	368	conserved hypothetical protein TIGR00048
SAG0459	171	VanZF domain protein
SAG0460	581	ABC transporter, ATP-binding/permease protein
SAG0461	579	ABC transporter, ATP-binding/permease protein
SAG0462	188	anthranilate synthase component II
SAG0463	179	BioY family protein
SAG0464	330	biotin synthetase
SAG0465	164	hypothetical protein
SAG0466	371	thiolase
SAG0467	409	AMP-binding enzyme domain protein
SAG0468	210	endonuclease III
SAG0469	131	type IV prepilin peptidase-related protein
SAG0470	69	conserved hypothetical protein
SAG0471	322	glucokinase
SAG0472	126	rhodanese-like family protein
SAG0473	613	elongation factor Tu family protein
SAG0474	81	conserved hypothetical protein
SAG0475	451	UDP-N-acetylmuramoylalanine--D-glutamate ligase
SAG0476	358	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide)

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
		pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
SAG0477	378	cell division protein DivIB, putative
SAG0478	429	cell division protein FtsA
SAG0479	426	cell division protein FtsZ
SAG0480	224	ylmE protein, putative
SAG0481	201	ylmF protein
SAG0482	84	YGGT family protein
SAG0483	262	ylmH protein
SAG0484	256	cell division protein DivIVA, putative
SAG0485	930	isoleucyl-tRNA synthetase
SAG0486	100	conserved hypothetical protein
SAG0487	151	MutT/nudix family protein
SAG0488	753	ATP-dependent Clp protease, ATP-binding subunit
SAG0489	34	hypothetical protein
SAG0490	76	conserved hypothetical protein
SAG0491	230	amino acid ABC transporter, permease protein
SAG0492	244	amino acid ABC transporter, ATP-binding protein
SAG0493	564	phosphoglucomutase/phosphomannomutase family protein
SAG0494	284	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
SAG0495	278	protein of unknown function
SAG0496	446	exodeoxyribonuclease VII, large subunit
SAG0497	71	exodeoxyribonuclease VII, small subunit
SAG0498	290	geranyltranstransferase, putative
SAG0499	275	hemolysin A
SAG0500	157	arginine repressor ArgR, putative
SAG0501	552	DNA repair protein RecN
SAG0502	278	DegV family protein
SAG0503	279	lipase/acylhydrolase
SAG0504	200	conserved hypothetical protein
SAG0505	91	DNA-binding protein HU
SAG0506	65	hypothetical protein
SAG0507	310	dihydroorotate dehydrogenase A
SAG0508	411	beta-lactam resistance factor
SAG0509	403	beta-lactam resistance factor
SAG0510	406	murM protein, putative
SAG0511	270	hydrolase, haloacid dehalogenase-like family
SAG0512	438	HD domain protein
SAG0513	128	conserved hypothetical protein
SAG0514	894	cation-transporting ATPase, E1-E2 family
SAG0515	286	conserved hypothetical protein
SAG0516	643	fructose-1,6-bisphosphatase, putative
SAG0517	374	iron-sulfur cluster-binding protein, putative
SAG0518	NA	peptide chain release factor 2, programmed frameshift
SAG0519	230	cell division ABC transporter, ATP-binding protein FtsE
SAG0520	309	cell division ABC transporter, permease protein FtsX
SAG0521	236	carboxymethylenebutenolidase-related protein
SAG0522	232	metallo-beta-lactamase superfamily protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0523	254	oxidoreductase, short chain dehydrogenase/reductase family
SAG0524	835	DNA polymerase III, epsilon subunit/ATP-dependent helicase DinG
SAG0525	397	aspartate aminotransferase
SAG0526	448	asparaginyl-tRNA synthetase
SAG0527	185	conserved hypothetical protein
SAG0528	327	inosine-uridine preferring nucleoside hydrolase
SAG0529	38	hypothetical protein
SAG0530	137	OsmC/Ohr family protein
SAG0531	296	conserved hypothetical protein
SAG0532	324	conserved hypothetical protein
SAG0533	303	conserved hypothetical protein
SAG0534	465	dipeptidase
SAG0535	506	zinc ABC transporter, zinc-binding adhesion liprotein
SAG0536	86	ribosomal protein L31
SAG0537	311	DHH family protein
SAG0538	340	adenosine deaminase, putative
SAG0539	147	flavodoxin
SAG0540	91	chorismate mutase, putative
SAG0541	398	voltage-gated chloride channel family protein
SAG0542	127	IS1381, transposase OrfA
SAG0543	129	IS1381, transposase OrfB
SAG0544	115	ribosomal protein L19
SAG0545	359	prophage LambdaSa1, site-specific recombinase, phage integrase family
SAG0546	67	conserved domain protein
SAG0547	185	hypothetical protein
SAG0548	265	prophage LambdaSa1, repressor protein, putative
SAG0549	47	hypothetical protein
SAG0550	74	conserved hypothetical protein
SAG0551	52	conserved hypothetical protein
SAG0552	62	hypothetical protein
SAG0553	268	hypothetical protein
SAG0554	63	prophage LambdaSa1, transcriptional regulator, Cro/CI family
SAG0555	249	prophage LambdaSa1, antirepressor, putative
SAG0556	47	hypothetical protein
SAG0557	76	hypothetical protein
SAG0558	74	hypothetical protein
SAG0559	286	conserved hypothetical protein
SAG0560	77	conserved hypothetical protein
SAG0561	46	hypothetical protein
SAG0562	84	hypothetical protein
SAG0563	53	hypothetical protein
SAG0564	160	conserved hypothetical protein
SAG0565	224	conserved domain protein
SAG0566	138	prophage LambdaSa1, single-strand binding protein
SAG0567	439	prophage LambdaSa1, reverse transcriptase/maturase family protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0568	67	conserved hypothetical protein
SAG0569	158	conserved hypothetical protein
SAG0570	115	hypothetical protein
SAG0571	43	hypothetical protein
SAG0572	138	conserved hypothetical protein
SAG0573	54	hypothetical protein
SAG0574	89	conserved hypothetical protein
SAG0575	110	hypothetical protein
SAG0576	43	hypothetical protein
SAG0577	177	conserved hypothetical protein
SAG0578	88	conserved hypothetical protein
SAG0579	142	conserved hypothetical protein
SAG0580	111	conserved hypothetical protein, truncation
SAG0581	118	conserved hypothetical protein
SAG0582	422	conserved hypothetical protein
SAG0583	406	conserved hypothetical protein
SAG0584	62	conserved hypothetical protein, truncation
SAG0585	471	conserved hypothetical protein
SAG0586	154	conserved hypothetical protein
SAG0587	300	prophage LambdaSa1, structural protein, putative
SAG0588	71	conserved hypothetical protein
SAG0589	143	conserved hypothetical protein
SAG0590	112	conserved hypothetical protein
SAG0591	78	conserved hypothetical protein
SAG0592	111	conserved hypothetical protein
SAG0593	185	prophage LambdaSa1, structural protein
SAG0594	81	conserved hypothetical protein
SAG0595	123	conserved hypothetical protein
SAG0596	670	prophage LambdaSa1, pblA protein, internal deletion
SAG0597	506	prophage LambdaSa1, minor structural protein, putative
SAG0598	1374	prophage LambdaSa1, N-acetylmuramoyl-L-alanine amidase, family 4
SAG0599	668	prophage LambdaSa1, minor structural protein, putative
SAG0600	109	hypothetical protein
SAG0601	70	hypothetical protein
SAG0602	100	conserved hypothetical protein
SAG0603	111	conserved hypothetical protein
SAG0604	239	prophage LambdaSa1, lysin, putative
SAG0605	323	conserved hypothetical protein
SAG0606	66	conserved hypothetical protein
SAG0607	56	conserved hypothetical protein
SAG0608	59	hypothetical protein
SAG0609	NA	prophage LambdaSa1, integrase, degenerate
SAG0610	134	conserved hypothetical protein
SAG0611	NA	transposase, degenerate
SAG0612	53	conserved hypothetical protein
SAG0613	425	transmembrane protein Vexp1
SAG0614	218	ABC transporter, ATP-binding protein Vexp2

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0615	458	transmembrane protein Vexp3
SAG0616	217	DNA-binding response regulator VncR
SAG0617	439	sensor histidine kinase VncS
SAG0618	195	transposase OrfB, IS3 family, truncation
SAG0619	66	conserved hypothetical protein
SAG0620	62	hypothetical protein
SAG0621	401	rod shape-determining protein RodA, putative□
SAG0622	186	hydrolase, haloacid dehalogenase-like family
SAG0623	650	DNA gyrase, B subunit
SAG0624	574	septation ring formation regulator EzrA, putative
SAG0625	213	phosphoserine phosphatase SerB
SAG0626	161	MutT/nudix family protein
SAG0627	151	conserved hypothetical protein
SAG0628	435	enolase
SAG0629	354	conserved domain protein
SAG0630	427	3-phosphoshikimate 1-carboxyvinyltransferase
SAG0631	170	shikimate kinase
SAG0632	457	psr protein
SAG0633	451	RNA methyltransferase, TrmA family
SAG0634	70	hypothetical protein
SAG0635	245	acid phosphatase, class B
SAG0636	172	conserved hypothetical protein
SAG0637	NA	transcriptional regulator, TetR family, putative, authentic frameshift
SAG0638	109	cell wall surface anchor family protein, truncation
SAG0639	273	transposase OrfB, IS3 family
SAG0640	91	transposase OrfA, IS3 family
SAG0641	NA	Tn5252, Orf 10 protein, degenerate
SAG0642	59	hypothetical protein
SAG0643	NA	chaperonin, 33 kDa, degenerate
SAG0644	402	transcriptional regulator, AraC family
SAG0645	554	cell wall surface anchor family protein
SAG0646	307	cell wall surface anchor family protein
SAG0647	305	sortase family protein
SAG0648	260	sortase family protein
SAG0649	890	cell wall surface anchor family protein, putative
SAG0650	189	sortase family protein
SAG0651	201	protein of unknown function
SAG0652	NA	Tn5252, Orf 28 protein, degenerate
SAG0653	NA	conserved hypothetical protein, degenerate
SAG0654	34	hypothetical protein
SAG0655	57	conserved hypothetical protein
SAG0656	36	hypothetical protein
SAG0657	89	hypothetical protein
SAG0658	383	lipoprotein, putative
SAG0659	330	ABC transporter, ATP-binding protein
SAG0660	272	membrane protein
SAG0661	261	conserved hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0662	101	cylX protein
SAG0663	282	cylD protein
SAG0664	240	cylG protein
SAG0665	101	acyl carrier protein AcpC
SAG0666	158	cylZ protein
SAG0667	309	cylA protein
SAG0668	292	cylB protein
SAG0669	667	cylE protein
SAG0670	317	cylF protein
SAG0671	731	cylI protein
SAG0672	403	cylJ protein
SAG0673	191	cylK protein
SAG0674	113	hypothetical protein
SAG0675	171	putative secreted protein
SAG0676	885	proteinase, putative
SAG0677	1062	hypothetical protein
SAG0678	NA	endopeptidase O, degenerate
SAG0679	343	protein of unknown function
SAG0680	339	protein of unknown function
SAG0681	353	conserved domain protein
SAG0682	409	permease, putative
SAG0683	NA	transmembrane protein Vexp3, putative, degenerate
SAG0684	223	ABC transporter, ATP-binding protein
SAG0685	472	conserved hypothetical protein
SAG0686	261	DNA-entry nuclease, putative
SAG0687	212	DedA family protein, putative
SAG0688	218	ABC transporter, ATP-binding protein
SAG0689	257	membrane protein, putative
SAG0690	272	conserved hypothetical protein
SAG0691	294	transcriptional regulator, LysR family
SAG0692	193	regulatory protein, putative
SAG0693	377	IS1548, transposase
SAG0694	173	regulatory protein, putative, truncation
SAG0695	330	D-lactate dehydrogenase
SAG0696	516	sodium:galactoside symporter family protein, putative
SAG0697	341	2-keto-3-deoxygluconate kinase
SAG0698	599	beta-glucuronidase
SAG0699	223	transcriptional regulator, GntR family
SAG0700	205	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase
SAG0701	466	glucuronate isomerase
SAG0702	348	mannonate dehydratase
SAG0703	279	D-mannonate oxidoreductase
SAG0704	270	hydrolase, haloacid dehalogenase-like family
SAG0705	596	glycosyl hydrolase, family 3
SAG0706	361	proline dipeptidase
SAG0707	334	transcriptional regulator, RegM family
SAG0708	488	alpha amylase family protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0709	332	glycosyl transferase, group 1 family protein
SAG0710	444	glycosyl transferase, group 1 family protein
SAG0711	647	threonyl-tRNA synthetase
SAG0712	234	DNA-binding response regulator
SAG0713	339	conserved hypothetical protein
SAG0714	188	conserved hypothetical protein
SAG0715	216	amino acid ABC transporter, permease protein
SAG0716	231	amino acid ABC transporter, permease protein
SAG0717	266	amino acid ABC transporter, amino acid-binding protein
SAG0718	251	amino acid ABC transporter, ATP-binding protein
SAG0719	236	DNA-binding response regulator
SAG0720	449	sensory box histidine kinase
SAG0721	269	metallo-beta-lactamase superfamily protein
SAG0722	122	conserved hypothetical protein
SAG0723	236	ribonuclease III
SAG0724	1179	chromosome segregation SMC protein
SAG0725	265	hydrolase, haloacid dehalogenase-like family
SAG0726	274	hydrolase, haloacid dehalogenase-like family
SAG0727	536	signal recognition particle-docking protein FtsY
SAG0728	270	ABC transporter, substrate-binding protein
SAG0729	300	ABC transporter, permease protein, putative
SAG0730	42	ABC transporter, ATP-binding protein
SAG0731	347	bacterial luciferase family protein
SAG0732	720	transcriptional accessory protein Tex, putative
SAG0733	142	conserved hypothetical protein
SAG0734	87	phage shock protein C, putative
SAG0735	44	hypothetical protein
SAG0736	311	HPr(Ser) kinase/phosphatase
SAG0737	257	prolipoprotein diacylglycerol transferase
SAG0738	132	conserved hypothetical protein
SAG0739	143	conserved hypothetical protein
SAG0740	91	conserved hypothetical protein
SAG0741	303	peptidase, U32 family, putative
SAG0742	428	peptidase, U32 family
SAG0743	70	conserved hypothetical protein
SAG0744	265	membrane protein, putative
SAG0745	446	Mn ²⁺ /Fe ²⁺ transporter, NRAMP family
SAG0746	369	riboflavin biosynthesis protein RibD
SAG0747	208	riboflavin synthase, alpha subunit
SAG0748	397	riboflavin biosynthesis protein RibA
SAG0749	156	riboflavin synthase, beta subunit
SAG0750	496	lysyl-tRNA synthetase
SAG0751	300	hydrolase, haloacid dehalogenase-like family
SAG0752	213	phosphoglycerate mutase family protein
SAG0753	157	ebsC family protein, putative
SAG0754	205	conserved domain protein
SAG0755	282	peptidase, U32 family
SAG0756	174	conserved hypothetical protein

Table 1: Complet list f GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0757	129	protein of unknown function/lipoprotein, putative
SAG0758	599	oligoendopeptidase F, putative
SAG0759	931	phosphoenolpyruvate carboxylase
SAG0760	377	IS1548, transposase
SAG0761	422	cell division protein, FtsW/RodA/SpoVE family
SAG0762	398	translation elongation factor Tu
SAG0763	252	triosephosphate isomerase
SAG0764	230	phosphoglycerate mutase family protein
SAG0765	681	penicillin-binding protein 2b
SAG0766	198	recombination protein RecR
SAG0767	348	D-alanine--D-alanine ligase
SAG0768	455	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-- D-alanyl-D-alanyl ligase
SAG0769	406	oxalate:formate antiporter
SAG0770	228	membrane protein, putative
SAG0771	512	cell wall surface anchor family protein
SAG0772	514	peptide chain release factor 3
SAG0773	126	conserved hypothetical protein
SAG0774	244	ABC transporter, ATP-binding protein
SAG0775	220	ABC transporter, permease protein
SAG0776	276	YaeC family protein, putative
SAG0777	528	ATP-dependent RNA helicase, DEAD/DEAH box family
SAG0778	88	conserved hypothetical protein
SAG0779	254	conserved hypothetical protein
SAG0780	246	acyltransferase family protein
SAG0781	217	competence protein CelsA
SAG0782	745	DNA internalization-related competence protein ComEC/Rec2
SAG0783	269	hydrolase, haloacid dehalogenase-like family
SAG0784	314	sugar-binding transcriptional regulator, LacI family
SAG0785	330	conserved hypothetical protein
SAG0786	242	conserved domain protein
SAG0787	345	DNA polymerase III, delta subunit, putative
SAG0788	202	superoxide dismutase, Fe-Mn
SAG0789	283	transcriptional antiterminator LicT
SAG0790	622	PTS system, beta-glucosides-specific IIBC components
SAG0791	475	6-phospho-beta-glucosidase
SAG0792	364	conserved hypothetical protein
SAG0793	380	glycerate kinase 2
SAG0794	418	permease, GntP family
SAG0795	354	conserved hypothetical protein
SAG0796	147	transcriptional regulator, MarR family
SAG0797	342	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
SAG0798	226	membrane protein, putative
SAG0799	233	glucosamine-6-phosphate isomerase
SAG0800	318	glutathione S-transferase family protein
SAG0801	239	ribosomal small subunit pseudouridine synthase A
SAG0802	38	hypothetical protein
SAG0803	383	major facilitator family protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0804	315	competence protein CoiA
SAG0805	601	oligoendopeptidase B
SAG0806	208	hydrolase, haloacid dehalogenase-like family
SAG0807	235	O-methyltransferase family protein
SAG0808	309	protease maturation protein, putative
SAG0809	161	conserved hypothetical protein
SAG0810	872	alanyl-tRNA synthetase
SAG0811	238	membrane protein, putative
SAG0812	272	glycosyl transferase, family 8
SAG0813	81	hypothetical protein
SAG0814	95	conserved hypothetical protein
SAG0815	71	transcriptional regulator, Cro/CI family
SAG0816	253	membrane protein, putative
SAG0817	187	membrane protein, putative
SAG0818	319	ribonucleoside-diphosphate reductase 2, beta subunit
SAG0819	719	ribonucleoside-diphosphate reductase 2, alpha subunit
SAG0820	74	ribonucleoside-diphosphate reductase 2, NrdH-redoxin
SAG0821	87	phosphocarrier protein HPr
SAG0822	577	phosphoenolpyruvate-protein phosphotransferase
SAG0823	475	glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent
SAG0824	417	polysaccharide deacetylase family protein
SAG0825	360	ATP-dependent RNA helicase, DEAD/DEAH box family
SAG0826	209	uridine kinase
SAG0827	165	conserved hypothetical protein
SAG0828	554	DNA polymerase III, gamma and tau subunits
SAG0829	64	conserved hypothetical protein
SAG0830	311	biotin--acetyl-CoA-carboxylase ligase
SAG0831	398	S-adenosylmethionine synthetase
SAG0832	753	protein of unknown function
SAG0833	181	hypothetical protein
SAG0834	42	hypothetical protein
SAG0835	188	conserved hypothetical protein
SAG0836	184	conserved hypothetical protein
SAG0837	428	ABC transporter, ATP-binding protein
SAG0838	233	hypothetical protein
SAG0839	226	transcriptional regulator, TenA family
SAG0840	265	phosphomethylpyrimidine kinase
SAG0841	256	hydroxyethylthiazole kinase
SAG0842	223	thiamine-phosphate pyrophosphorylase
SAG0843	419	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
SAG0844	184	acetyltransferase, GNAT family
SAG0845	427	CBS domain protein
SAG0846	286	methionine aminopeptidase, type I
SAG0847	306	ribonuclease BN, putative
SAG0848	151	GtrA family protein
SAG0849	169	conserved hypothetical protein
SAG0850	652	DNA ligase, NAD-dependent
SAG0851	339	bmrU protein, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0852	766	pullulanase, putative
SAG0853	622	1,4- α -glucan branching enzyme
SAG0854	379	glucose-1-phosphate adenylyltransferase
SAG0855	NA	glycogen biosynthesis protein GlgD, authentic frameshift
SAG0856	476	glycogen synthase
SAG0857	66	ATP synthase F0, C subunit
SAG0858	238	ATP synthase F0, A subunit
SAG0859	165	ATP synthase F0, B subunit
SAG0860	178	ATP synthase F1, delta subunit
SAG0861	501	ATP synthase F1, alpha subunit
SAG0862	293	ATP synthase F1, gamma subunit
SAG0863	468	ATP synthase F1, beta subunit
SAG0864	137	ATP synthase F1, epsilon subunit
SAG0865	76	conserved hypothetical protein
SAG0866	423	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
SAG0867	63	conserved hypothetical protein
SAG0868	285	DNA-entry nuclease
SAG0869	346	phenylalanyl-tRNA synthetase, alpha subunit
SAG0870	173	acetyltransferase, GNAT family
SAG0871	801	phenylalanyl-tRNA synthetase, beta subunit
SAG0872	300	conserved hypothetical protein
SAG0873	1077	exonuclease RexB
SAG0874	1207	exonuclease RexA
SAG0875	305	magnesium transporter, CorA family, putative
SAG0876	458	tRNA modification GTPase TrmE
SAG0877	636	ABC transporter, ATP-binding protein
SAG0878	322	acetoin dehydrogenase, thymine PPi dependent, E1 component, alpha subunit
SAG0879	332	acetoin dehydrogenase, thymine PPi dependent, E1 component, beta subunit
SAG0880	462	acetoin dehydrogenase, thymine PPi dependent, E2 component, dihydrolipoamide acetyltransferase
SAG0881	585	acetoin dehydrogenase, thymine PPi dependent, E3 component, dihydrolipoamide dehydrogenase
SAG0882	329	lipoate-protein ligase A
SAG0883	261	cobyric acid synthase, putative
SAG0884	447	mur ligase family protein
SAG0885	283	conserved hypothetical protein TIGR00159
SAG0886	319	protein of unknown function
SAG0887	450	phosphoglucomutase/phosphomannomutase family protein
SAG0888	123	conserved hypothetical protein
SAG0889	126	conserved hypothetical protein
SAG0890	376	oxygen-independent coproporphyrinogen III oxidase, putative
SAG0891	245	conserved hypothetical protein
SAG0892	256	hydrolase, haloacid dehalogenase-like family
SAG0893	218	conserved hypothetical protein
SAG0894	1370	protein of unknown function
SAG0895	289	lipoyl-binding domain protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0896	108	oxidoreductase, putative
SAG0897	221	conserved hypothetical protein
SAG0898	83	hypothetical protein
SAG0899	57	hypothetical protein
SAG0900	56	hypothetical protein
SAG0901	127	hypothetical protein
SAG0902	45	hypothetical protein
SAG0903	44	hypothetical protein
SAG0904	56	hypothetical protein
SAG0905	138	nucleoside diphosphate kinase
SAG0906	610	GTP-binding protein LepA
SAG0907	877	protein of unknown function/lipoprotein, putative
SAG0908	203	HD domain protein
SAG0909	154	acetyltransferase, GNAT family
SAG0910	144	PilB-related protein
SAG0911	930	cation-transporting ATPase, E1-E2 family
SAG0912	367	nucleoside diphosphate kinase domain protein
SAG0913	212	chloramphenicol acetyltransferase
SAG0914	203	conserved hypothetical protein
SAG0915	405	Tn916, transposase
SAG0916	67	Tn916, excisionase
SAG0917	83	Tn916, hypothetical protein
SAG0918	76	Tn916, hypothetical protein
SAG0919	157	Tn916, hypothetical protein
SAG0920	23	Tn916, hypothetical protein
SAG0921	117	Tn916, transcriptional regulator, putative
SAG0922	61	Tn916, hypothetical protein
SAG0923	639	Tn916, tetracycline resistance protein
SAG0924	28	Tn916, tetM leader peptide
SAG0925	310	Tn916, hypothetical protein
SAG0926	333	Tn916, NLP/P60 family protein
SAG0927	725	membrane protein, putative
SAG0928	NA	Tn916, hypothetical protein, authentic frameshift
SAG0929	168	Tn916, hypothetical protein
SAG0930	165	Tn916, hypothetical protein
SAG0931	73	Tn916, hypothetical protein
SAG0932	401	Tn916, transcriptional regulator, putative
SAG0933	461	Tn916, FtsK/SpoIIIE family protein
SAG0934	128	Tn916, hypothetical protein
SAG0935	104	Tn916, hypothetical protein
SAG0936	39	Tn916, hypothetical protein
SAG0937	NA	ABC transporter, ATP-binding protein, authentic frameshift
SAG0938	122	transcriptional regulator, GntR family
SAG0939	1034	DNA polymerase III, alpha subunit
SAG0940	340	6-phosphofructokinase
SAG0941	500	pyruvate kinase
SAG0942	185	signal peptidase I, putative
SAG0943	47	hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0944	604	glucosamine--fructose-6-phosphate aminotransferase, isomerizing
SAG0945	377	IS1548, transposase
SAG0946	109	phnA protein
SAG0947	213	amino acid ABC transporter, permease protein
SAG0948	209	amino acid ABC transporter, ATP-binding protein
SAG0949	276	amino acid ABC transporter, amino acid-binding protein
SAG0950	82	ribosomal protein S20
SAG0951	306	pantothenate kinase
SAG0952	196	conserved hypothetical protein
SAG0953	129	cytidine deaminase
SAG0954	349	protein of unknown function/lipoprotein, putative
SAG0955	511	sugar ABC transporter, ATP-binding protein
SAG0956	353	sugar ABC transporter, permease protein, putative
SAG0957	318	sugar ABC transporter, permease protein, putative
SAG0958	456	NADH oxidase
SAG0959	329	L-lactate dehydrogenase
SAG0960	819	DNA gyrase, A subunit
SAG0961	247	sortase SrtA
SAG0962	137	glyoxylase family protein
SAG0963	320	conserved hypothetical protein
SAG0964	375	Na ⁺ /H ⁺ exchanger family protein
SAG0965	127	IS1381, transposase OrfA
SAG0966	129	IS1381, transposase OrfB
SAG0967	520	GMP synthase
SAG0968	232	transcriptional regulator, GntR family
SAG0969	444	gid protein
SAG0970	247	acetyltransferase, GNAT family
SAG0971	282	protein of unknown function/lipoprotein, putative
SAG0972	NA	conserved hypothetical protein, authentic frameshift
SAG0973	320	nisin-resistance protein, putative
SAG0974	250	ABC transporter, ATP-binding protein
SAG0975	651	ABC transporter, permease protein, putative
SAG0976	222	DNA-binding response regulator
SAG0977	312	sensor histidine kinase
SAG0978	356	site-specific recombinase, phage integrase family
SAG0979	553	ABC transporter, substrate-binding protein
SAG0980	257	conserved hypothetical protein
SAG0981	228	satD protein
SAG0982	521	signal recognition particle protein Ffh
SAG0983	110	conserved hypothetical protein
SAG0984	437	sensor histidine kinase CiaH
SAG0985	226	DNA-binding response regulator CiaR
SAG0986	849	aminopeptidase N
SAG0987	217	phosphate transport system regulatory protein PhoU
SAG0988	252	phosphate ABC transporter, ATP-binding protein PstB, putative
SAG0989	267	phosphate ABC transporter, ATP-binding protein PstB, putative
SAG0990	295	phosphate ABC transporter, permease protein PstA, putative
SAG0991	305	phosphate ABC transporter, permease protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG0992	286	phosphate ABC transporter, phosphate-binding protein
SAG0993	436	NOL1/NOP2/sun family protein
SAG0994	254	inositol monophosphatase family protein
SAG0995	93	conserved hypothetical protein
SAG0996	137	conserved hypothetical protein
SAG0997	310	macrolide-efflux protein mreA/riboflavin biosynthesis protein RibF
SAG0998	294	tRNA pseudouridine synthase B
SAG0999	143	acetyltransferase, GNAT family
SAG1000	423	conserved hypothetical protein
SAG1001	196	conserved hypothetical protein
SAG1002	292	protease, putative
SAG1003	876	permease, putative
SAG1004	233	ABC transporter, ATP-binding protein
SAG1005	706	DNA topoisomerase I
SAG1006	280	DprA/SMF protein, putative DNA processing factor
SAG1007	342	iron-compound ABC transporter, iron-compound-binding protein
SAG1008	253	iron compound ABC transporter, ATP-binding protein
SAG1009	324	iron compound ABC transporter, permease protein
SAG1010	320	iron compound ABC transporter, permease protein
SAG1011	182	acetyltransferase, CysE/LacA/LpxA/NodL family
SAG1012	253	ribonuclease HII
SAG1013	283	GTP-binding protein
SAG1014	190	conserved hypothetical protein
SAG1015	494	carbon starvation protein CstA, putative
SAG1016	244	response regulator
SAG1017	579	sensor histidine kinase, putative
SAG1018	40	lipoprotein, putative
SAG1019	39	hypothetical protein
SAG1020	227	lipoprotein, putative
SAG1021	107	hypothetical protein
SAG1022	177	hypothetical protein
SAG1023	48	hypothetical protein
SAG1024	183	lipoprotein, putative
SAG1025	149	hypothetical protein
SAG1026	NA	immunogenic secreted protein, degenerate
SAG1027	84	conserved hypothetical protein
SAG1028	196	hypothetical protein
SAG1029	101	hypothetical protein
SAG1030	304	protein of unknown function
SAG1031	120	conserved domain protein
SAG1032	85	conserved hypothetical protein
SAG1033	1309	FtsK/SpoIIIE family protein
SAG1034	55	hypothetical protein
SAG1035	424	conserved hypothetical protein
SAG1036	80	conserved hypothetical protein
SAG1037	157	hypothetical protein
SAG1038	1003	phage infection protein, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1039	96	conserved hypothetical protein
SAG1040	260	conserved domain protein
SAG1041	107	hypothetical protein
SAG1042	1060	carbamoyl-phosphate synthase, large subunit
SAG1043	358	carbamoyl-phosphate synthase, small subunit
SAG1044	307	aspartate carbamoyltransferase
SAG1045	430	dihydroorotase, multifunctional complex type
SAG1046	209	orotate phosphoribosyltransferase
SAG1047	233	orotidine 5'-phosphate decarboxylase
SAG1048	410	membrane protein, putative
SAG1049	513	ABC transporter, ATP-binding protein
SAG1050	112	ribonucleotide reductase, truncation
SAG1051	358	aspartate-semialdehyde dehydrogenase
SAG1052	47	cell wall surface anchor family protein, putative
SAG1053	30	hypothetical protein
SAG1054	531	cardiolipin synthetase
SAG1055	556	formate--tetrahydrofolate ligase
SAG1056	339	lipoate-protein ligase A
SAG1057	292	conserved hypothetical protein
SAG1058	272	conserved hypothetical protein
SAG1059	110	glycine cleavage system H protein, putative
SAG1060	328	bacterial luciferase family protein
SAG1061	399	oxidoreductase, FMN-binding
SAG1062	282	lipoate-protein ligase A family protein
SAG1063	228	flavoprotein-related protein
SAG1064	180	flavoprotein family protein
SAG1065	190	membrane protein, putative
SAG1066	572	phosphoglucomutase
SAG1067	178	IS861, transposase OrfA
SAG1068	277	IS861, transposase OrfB
SAG1069	65	hypothetical protein
SAG1070	577	ABC transporter, ATP-binding/permease protein
SAG1071	573	ABC transporter, ATP-binding/permease protein
SAG1072	200	conserved hypothetical protein
SAG1073	325	conserved hypothetical protein
SAG1074	418	serine hydroxymethyltransferase
SAG1075	183	Sua5/YciO/YrdC/YwlC family protein
SAG1076	276	modification methylase, HemK family
SAG1077	359	peptide chain release factor 1
SAG1078	189	thymidine kinases
SAG1079	60	4-oxalocrotonate tautomerase
SAG1080	47	hypothetical protein
SAG1081	312	ApbE family protein
SAG1082	200	conserved hypothetical protein
SAG1083	411	conserved hypothetical protein
SAG1084	262	formate/nitrite transporter family protein
SAG1085	424	xanthine permease
SAG1086	193	xanthine phosphoribosyltransferase

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1087	327	guanosine monophosphate reductase
SAG1088	446	drug resistance transporter, EmrB/QacA family, putative
SAG1089	230	conserved hypothetical protein
SAG1090	666	potassium uptake protein, putative
SAG1091	216	oxidoreductase, short chain dehydrogenase/reductase family
SAG1092	330	phosphate acetyltransferase
SAG1093	294	ribosomal large subunit pseudouridine synthase, RluD subfamily
SAG1094	278	conserved hypothetical protein
SAG1095	223	GTP pyrophosphokinase family protein
SAG1096	190	conserved hypothetical protein
SAG1097	324	ribose-phosphate pyrophosphokinase
SAG1098	371	cysteine desulphurase
SAG1099	115	conserved hypothetical protein
SAG1100	210	conserved hypothetical protein
SAG1101	226	DNA repair protein RadC
SAG1102	377	membrane protein, putative
SAG1103	478	6-phospho-beta-glucosidase
SAG1104	204	platelet activating factor, putative
SAG1105	273	hydrolase, haloacid dehalogenase-like family
SAG1106	309	transcriptional regulator, AraC family, putative
SAG1107	510	voltage-gated chloride channel family protein
SAG1108	357	spermidine/putrescine ABC transporter, spermidine/putrescine-binding protein
SAG1109	258	spermidine/putrescine ABC transporter, permease protein
SAG1110	264	spermidine/putrescine ABC transporter, permease protein
SAG1111	384	spermidine/putrescine ABC transporter, ATP-binding protein
SAG1112	300	UDP-N-acetylenolpyruvoylglucosamine reductase
SAG1113	162	2-amino-4-hydroxy-6-hydroxymethyl dihydropteridine pyrophosphokinase
SAG1114	120	dihydroneopterin aldolase
SAG1115	267	dihydropteroate synthase
SAG1116	187	GTP cyclohydrolase I
SAG1117	420	folylpolyglutamate synthase
SAG1118	295	rarD protein
SAG1119	288	homoserine kinase
SAG1120	427	homoserine dehydrogenase
SAG1121	295	polysaccharide deacetylase family protein
SAG1122	515	transporter, BCCT family protein
SAG1123	34	hypothetical protein
SAG1124	458	aldehyde dehydrogenase family protein
SAG1125	335	membrane protein, putative
SAG1126	228	protein of unknown function
SAG1127	446	conserved domain protein
SAG1128	65	transcriptional regulator, Cro/CI family
SAG1129	36	hypothetical protein
SAG1130	49	hypothetical protein
SAG1131	164	thiol peroxidase
SAG1132	219	conserved hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1133	254	conserved hypothetical protein
SAG1134	213	transcriptional regulator, GntR family/potassium uptake protein, TrkA family
SAG1135	183	gls24 protein, putative
SAG1136	65	conserved hypothetical protein
SAG1137	180	gls24 protein, putative
SAG1138	64	conserved hypothetical protein
SAG1139	193	conserved hypothetical protein
SAG1140	82	conserved hypothetical protein
SAG1141	112	conserved hypothetical protein
SAG1142	759	ATP-dependent DNA helicase PcrA
SAG1143	128	conserved hypothetical protein
SAG1144	441	uracil permease
SAG1145	448	sodium:alanine symporter family protein
SAG1146	411	cation efflux family protein
SAG1147	130	conserved hypothetical protein
SAG1148	231	membrane protein, putative
SAG1149	207	lipoprotein, putative
SAG1150	400	ribosomal protein S1
SAG1151	76	conserved hypothetical protein
SAG1152	340	branched-chain amino acid aminotransferase
SAG1153	819	DNA topoisomerase IV, A subunit
SAG1154	653	DNA topoisomerase IV, B subunit
SAG1155	212	membrane protein, putative
SAG1156	217	uracil-DNA glycosylase
SAG1157	161	conserved hypothetical protein
SAG1158	413	CMP-N-acetylneuraminic acid synthetase NeuA
SAG1159	209	neuD protein
SAG1160	384	UDP-N-acetylglucosamine-2-epimerase NeuC
SAG1161	341	N-acetyl neuramic acid synthetase NeuB
SAG1162	466	polysaccharide biosynthesis protein CpsL
SAG1163	318	polysaccharide biosynthesis protein CpsK(V)
SAG1164	321	glycosyl transferase CpsJ(V)
SAG1165	327	glycosyl transferase CpsO(V)
SAG1166	295	glycosyl transferase CpsN(V)
SAG1167	241	polysaccharide biosynthesis protein CpsM(V)
SAG1168	364	polysaccharide biosynthesis protein cpsH(V)
SAG1169	163	glycosyl transferase CpsG(V)
SAG1170	149	polysaccharide biosynthesis protein CpsF
SAG1171	462	glycosyl transferase CpsE
SAG1172	229	cpsD protein
SAG1173	230	cpsC protein
SAG1174	243	capsular polysaccharide biosynthesis protein CpsB
SAG1175	485	capsular polysaccharide biosynthesis protein CpsA
SAG1176	290	transcriptional regulator, LysR family, putative
SAG1177	255	conserved hypothetical protein
SAG1178	236	purine nucleoside phosphorylase
SAG1179	418	voltage-gated chloride channel family protein, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1180	269	purine nucleoside phosphorylase
SAG1181	135	arsenate reductase
SAG1182	403	phosphopentomutase
SAG1183	223	ribose 5-phosphate isomerase
SAG1184	236	conserved hypothetical protein
SAG1185	262	tributylin esterase
SAG1186	553	metallo-beta-lactamase superfamily protein
SAG1187	253	ABC transporter, ATP-binding protein
SAG1188	287	ABC transporter, permease protein
SAG1189	334	conserved hypothetical protein
SAG1190	551	adherence and virulence protein A
SAG1191	239	alpha-acetolactate decarboxylase
SAG1192	560	acetolactate synthase, catabolic
SAG1193	408	TPR domain protein
SAG1194	396	membrane protein, putative
SAG1195	153	MutT/nudix family protein
SAG1196	160	mutator MutT protein
SAG1197	1072	hyaluronidase
SAG1198	348	dTDP-glucose 4,6-dehydratase
SAG1199	197	dTDP-4-dehydrorhamnose 3,5-epimerase
SAG1200	289	glucose-1-phosphate thymidyltransferase
SAG1201	367	iminodiacetate oxidase, putative
SAG1202	262	conserved hypothetical protein TIGR00486
SAG1203	227	conserved hypothetical protein
SAG1204	226	DNA replication protein DnaD, putative
SAG1205	172	adenine phosphoribosyltransferase
SAG1206	854	conserved domain protein
SAG1207	32	hypothetical protein
SAG1208	732	single-stranded-DNA-specific exonuclease RecJ
SAG1209	253	oxidoreductase, short chain dehydrogenase/reductase family
SAG1210	309	metallo-beta-lactamase superfamily protein
SAG1211	215	conserved hypothetical protein
SAG1212	412	GTP-binding protein HflX
SAG1213	296	tRNA delta(2)-isopentenylpyrophosphate transferase
SAG1214	58	hypothetical protein
SAG1215	305	exfoliative toxin A, putative
SAG1216	1252	pullulanase, putative
SAG1217	NA	conserved hypothetical protein, authentic frameshift
SAG1218	194	conserved hypothetical protein
SAG1219	468	peptidase, M20/M25/M40 family
SAG1220	200	nitroreductase family protein
SAG1221	NA	glycerophosphoryl diester phosphodiesterase, putative, authentic point mutation
SAG1222	593	excinuclease ABC, C subunit
SAG1223	255	conserved hypothetical protein
SAG1224	446	MATE efflux family protein
SAG1225	136	conserved hypothetical protein
SAG1226	165	conserved hypothetical protein

Table 1: Comple list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1227	198	protein of unknown function
SAG1228	96	ISSdy1, transposase OrfA
SAG1229	259	ISSdy1, transposase OrfB
SAG1230	96	conserved hypothetical protein
SAG1231	NA	transposase OrfB, IS3 family, degenerate
SAG1232	77	transposase OrfB, IS3 family, truncation
SAG1233	822	streptococcal histidine triad family protein
SAG1234	306	laminin-binding surface protein
SAG1235	425	GBSi1, group II intron, maturase
SAG1236	NA	C5a peptidase, authentic frameshift
SAG1237	444	hypothetical protein
SAG1238	202	hypothetical protein
SAG1239	76	conserved hypothetical protein
SAG1240	125	conserved hypothetical protein, truncation
SAG1241	91	transposase OrfA, IS3 family
SAG1242	67	transposase OrfB, IS3 family, truncation
SAG1243	96	ISSdy1, transposase OrfA
SAG1244	259	ISSdy1, transposase OrfB
SAG1245	38	hypothetical protein
SAG1246	389	hypothetical protein
SAG1247	399	site-specific recombinase, phage integrase family
SAG1248	75	conserved hypothetical protein
SAG1249	74	transcriptional regulator, Cro/CI family
SAG1250	621	Tn5252, relaxase
SAG1251	121	Tn5252, Orf 9 protein
SAG1252	120	Tn5252, Orf 10 protein
SAG1253	435	transposase, ISL3 family
SAG1254	546	mercuric reductase
SAG1255	130	mercuric resistance operon regulatory protein MerR
SAG1256	142	IS861, transposase OrfB, truncation
SAG1257	709	cation-transporting ATPase, E1-E2 family
SAG1258	122	cadmium efflux system accessory protein
SAG1259	99	conserved hypothetical protein
SAG1260	262	hypothetical protein
SAG1261	198	conserved hypothetical protein
SAG1262	695	cation-transporting ATPase, E1-E2 family
SAG1263	NA	conserved domain protein, authentic frameshift
SAG1264	148	transcriptional repressor CopY, putative
SAG1265	206	cadmium resistance transporter, putative
SAG1266	152	hypothetical protein
SAG1267	108	hypothetical protein
SAG1268	230	repressor protein, putative
SAG1269	44	hypothetical protein
SAG1270	471	ImpB/MucB/SamB family protein
SAG1271	116	conserved hypothetical protein
SAG1272	102	conserved hypothetical protein
SAG1273	118	conserved hypothetical protein
SAG1274	129	conserved hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1275	75	hypothetical protein
SAG1276	358	conserved hypothetical protein
SAG1277	163	hypothetical protein
SAG1278	96	hypothetical protein
SAG1279	99	conserved domain protein
SAG1280	2274	SNF2 family protein
SAG1281	183	hypothetical protein
SAG1282	63	calcium-binding protein, putative
SAG1283	1631	agglutinin receptor
SAG1284	196	abortive infection protein AbiGI
SAG1285	281	abortive infection protein AbiGII
SAG1286	933	Tn5252, Orf28
SAG1287	776	Tn5252, Orf26
SAG1288	NA	Tn5252, Orf25, degenerate
SAG1289	284	Tn5252, Orf23
SAG1290	80	hypothetical protein
SAG1291	605	Tn5252, Orf 21 protein, internal deletion
SAG1292	162	hypothetical protein
SAG1293	194	protease, putative
SAG1294	77	conserved hypothetical protein
SAG1295	127	conserved hypothetical protein
SAG1296	142	conserved hypothetical protein
SAG1297	451	C-5 cytosine-specific DNA methylase
SAG1298	31	hypothetical protein
SAG1299	272	conserved hypothetical protein
SAG1300	57	conserved hypothetical protein
SAG1301	121	ribosomal protein L7/L12
SAG1302	166	ribosomal protein L10
SAG1303	702	ATP-dependent Clp protease, ATP-binding subunit
SAG1304	32	hypothetical protein
SAG1305	314	homocysteine S-methyltransferase MmuM, putative
SAG1306	458	amino acid permease
SAG1307	216	hypothetical protein
SAG1308	167	hypothetical protein
SAG1309	30	hypothetical protein
SAG1310	182	transcriptional regulator, TetR family
SAG1311	198	GTP-binding protein
SAG1312	408	ATP-dependent Clp protease, ATP-binding subunit ClpX
SAG1313	56	conserved hypothetical protein
SAG1314	164	dihydrofolate reductase
SAG1315	279	thymidylate synthase
SAG1316	390	HMG-CoA synthase
SAG1317	427	3-hydroxy-3-methylglutaryl-CoA reductase
SAG1318	149	conserved hypothetical protein
SAG1319	214	hemolysin III, putative
SAG1320	304	conserved hypothetical protein TIGR00147
SAG1321	284	glutathione S-transferase family protein, putative
SAG1322	72	conserved domain protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1323	331	isopentenyl-diphosphate delta-isomerase
SAG1324	330	phosphomevalonate kinase
SAG1325	314	diphosphomevalonate decarboxylase
SAG1326	292	mevalonate kinase, putative
SAG1327	409	sensor histidine kinase
SAG1328	228	DNA-binding response regulator
SAG1329	208	GTP pyrophosphokinase family protein
SAG1330	68	hypothetical protein
SAG1331	979	R5 protein
SAG1332	146	transcriptional regulator, MarR family, putative
SAG1333	690	5'-nucleotidase family protein
SAG1334	136	polypeptide deformylase, putative
SAG1335	449	NADP-specific glutamate dehydrogenase
SAG1336	169	membrane protein, putative
SAG1337	589	ABC transporter, ATP-binding/permease protein
SAG1338	579	ABC transporter, ATP-binding/permease protein
SAG1339	157	acetyltransferase, GNAT family
SAG1340	622	ABC transporter, ATP-binding protein
SAG1341	402	polyA polymerase family protein
SAG1342	282	DegV family protein
SAG1343	126	protein of unknown function
SAG1344	177	hypothetical protein
SAG1345	164	conserved hypothetical protein
SAG1346	654	PTS system, fructose specific IIABC components
SAG1347	303	1-phosphofructokinase
SAG1348	247	lactose phosphotransferase system repressor
SAG1349	411	beta-lactam resistance factor
SAG1350	544	surface antigen-related protein
SAG1351	307	2-dehydropantoate 2-reductase, putative
SAG1352	356	regulatory protein, putative
SAG1353	330	pyridine nucleotide-disulphide oxidoreductase family protein
SAG1354	251	tRNA (guanine-N1)-methyltransferase
SAG1355	172	16S rRNA processing protein RimM
SAG1356	503	transcriptional regulator, RofA family
SAG1357	80	KH domain protein
SAG1358	90	ribosomal protein S16
SAG1359	415	permease, putative
SAG1360	236	ABC transporter, ATP-binding protein
SAG1361	414	conserved hypothetical protein
SAG1362	532	carbamoyl-phosphate synthase, large subunit, putative
SAG1363	356	carbamoyl-phosphate synthase, small subunit
SAG1364	173	pyrimidine operon regulatory protein
SAG1365	296	ribosomal large subunit pseudouridine synthase, RluD subfamily
SAG1366	154	lipoprotein signal peptidase
SAG1367	301	transcriptional regulator, LysR family
SAG1368	94	ribosomal protein L27
SAG1369	112	conserved hypothetical protein
SAG1370	104	ribosomal protein L21

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1371	392	conserved hypothetical protein
SAG1372	404	thiamine biosynthesis protein ThiI
SAG1373	381	cysteine desulphurase
SAG1374	150	conserved hypothetical protein
SAG1375	449	glutathione reductase
SAG1376	111	conserved hypothetical protein
SAG1377	388	chorismate synthase
SAG1378	355	3-dehydroquinate synthase
SAG1379	225	3-dehydroquinate dehydratase
SAG1380	385	conserved hypothetical protein
SAG1381	714	sulfatase
SAG1382	119	ribosomal protein L20
SAG1383	66	ribosomal protein L35
SAG1384	176	translation initiation factor IF-3
SAG1385	227	cytidylate kinase
SAG1386	174	conserved hypothetical protein
SAG1387	65	ferredoxin, 4Fe-4S
SAG1388	163	conserved hypothetical protein
SAG1389	406	peptidase T
SAG1390	544	polysaccharide biosynthesis protein, putative
SAG1391	484	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase
SAG1392	264	iron compound ABC transporter, ATP-binding protein
SAG1393	310	iron compound ABC transporter, substrate-binding protein
SAG1394	341	iron compound ABC transporter, permease protein
SAG1395	333	iron compound ABC transporter, permease protein
SAG1396	217	conserved hypothetical protein
SAG1397	311	inorganic pyrophosphatase, manganese-dependent
SAG1398	262	pyruvate formate-lyase-activating enzyme
SAG1399	444	CBS domain protein
SAG1400	188	conserved hypothetical protein
SAG1401	311	conserved hypothetical protein TIGR01212
SAG1402	213	PAP2 family protein
SAG1403	194	membrane protein, putative
SAG1404	308	cell wall surface anchor family protein
SAG1405	294	sortase family protein
SAG1406	293	sortase family protein
SAG1407	705	cell wall surface anchor family protein
SAG1408	901	cell wall surface anchor family protein
SAG1409	NA	rogB protein, authentic frameshift
SAG1410	379	glycosyl transferase, group 1 family protein
SAG1411	282	glycosyl transferase, group 2 family protein
SAG1412	474	polysaccharide biosynthesis protein
SAG1413	454	membrane protein, putative
SAG1414	308	glycosyl transferase, group 2 family protein
SAG1415	311	glycosyl transferase, group 2 family protein
SAG1416	352	nucleotide sugar dehydratase, putative
SAG1417	240	nucleotidyl transferase, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1418	274	polysaccharide biosynthesis protein, putative
SAG1419	577	lipoprotein, putative
SAG1420	117	conserved hypothetical protein
SAG1421	243	glycosyl transferase, group 2 family protein
SAG1422	313	glycosyl transferase, group 2 family protein
SAG1423	384	glycosyl transferase, putative
SAG1424	284	dTDP-4-dehydrorhamnose reductase
SAG1425	113	conserved hypothetical protein
SAG1426	369	RNA polymerase sigma-70 factor
SAG1427	602	DNA primase
SAG1428	125	large conductance mechanosensitive channel protein
SAG1429	58	ribosomal protein S21
SAG1430	167	conserved hypothetical protein
SAG1431	268	amino acid ABC transporter, amino acid-binding protein
SAG1432	347	ammonium transporter family protein
SAG1433	375	conserved hypothetical protein
SAG1434	328	rhodanese family protein
SAG1435	101	conserved hypothetical protein
SAG1436	457	glycerol-3-phosphate transporter, putative
SAG1437	55	hypothetical protein
SAG1438	754	glycogen phosphorylase
SAG1439	498	4-alpha-glucanotransferase
SAG1440	342	maltose operon repressor MalR, putative
SAG1441	415	maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein
SAG1442	456	maltose ABC transporter, permease protein
SAG1443	278	maltose ABC transporter, permease protein
SAG1444	490	proton/peptide symporter family protein
SAG1445	NA	MutT/nudix family protein, authentic frameshift
SAG1446	62	hypothetical protein
SAG1447	441	conserved hypothetical protein
SAG1448	502	glycosyl transferase, group 1 family protein
SAG1449	795	preprotein translocase SecA subunit, putative
SAG1450	330	conserved domain protein
SAG1451	494	conserved hypothetical protein
SAG1452	514	conserved hypothetical protein
SAG1453	409	preprotein translocase SecY family protein
SAG1454	398	glycosyl transferase, putative
SAG1455	295	glycosyl transferase, group 2 family protein
SAG1456	NA	glycosyl transferase, family 8, degenerate
SAG1457	129	IS1381, transposase OrfB
SAG1458	127	IS1381, transposase OrfA
SAG1459	413	glycosyl transferase family 8
SAG1460	401	glycosyl transferase, family 8
SAG1461	335	conserved hypothetical protein
SAG1462	970	cell wall surface anchor family protein
SAG1463	NA	transcriptional regulator, RofA family, authentic point mutation
SAG1464	663	excinuclease ABC, B subunit

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1465	306	protease, putative
SAG1466	727	glutamine ABC transporter, glutamine-binding protein/permease protein
SAG1467	246	glutamine ABC transporter, ATP-binding protein, GlnQ putative
SAG1468	116	conserved hypothetical protein
SAG1469	52	conserved hypothetical protein
SAG1470	437	GTP-binding protein, GTP1/Obg family
SAG1471	42	conserved hypothetical protein
SAG1472	413	aminopeptidase PepS
SAG1473	192	cell wall surface anchor family protein
SAG1474	680	amidase family protein
SAG1475	240	ribosomal small subunit pseudouridine synthase A
SAG1476	280	oxidoreductase, aldo/keto reductase family
SAG1477	224	nitroreductase family protein
SAG1478	130	lactoylglutathione lyase
SAG1479	308	glycosyl transferase, group 2 family protein
SAG1480	462	amino acid permease
SAG1481	155	SsrA-binding protein
SAG1482	801	exoribonuclease, VacB/Rnb family
SAG1483	78	preprotein translocase, SecG subunit
SAG1484	48	ribosomal protein L33
SAG1485	389	multi-drug resistance protein
SAG1486	548	membrane protein, putative
SAG1487	233	ABC transporter, ATP binding protein
SAG1488	195	dephospho-CoA kinase
SAG1489	273	formamidopyrimidine-DNA glycosylase
SAG1490	282	transcriptional regulator, MutR family
SAG1491	530	hypothetical protein
SAG1492	58	hypothetical protein
SAG1493	66	hypothetical protein
SAG1494	32	hypothetical protein
SAG1495	81	CAAX amino terminal protease family protein
SAG1496	110	hypothetical protein
SAG1497	37	hypothetical protein
SAG1498	133	hypothetical protein
SAG1499	299	GTP-binding protein Era
SAG1500	132	diacylglycerol kinase
SAG1501	161	conserved hypothetical protein TIGR00043
SAG1502	268	tetracenomycin polyketide synthesis O-methyltransferase TcmP, putative
SAG1503	39	hypothetical protein
SAG1504	38	hypothetical protein
SAG1505	158	MutT/nudix family protein
SAG1506	267	hypothetical protein
SAG1507	345	PhoH family protein
SAG1508	590	67 kDa Myosin-crossreactive streptococcal antigen
SAG1509	71	conserved hypothetical protein
SAG1510	169	peptide methionine sulfoxide reductase

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1511	284	conserved hypothetical protein
SAG1512	185	ribosome recycling factor
SAG1513	242	uridylate kinase
SAG1514	226	peptide ABC transporter, ATP-binding protein
SAG1515	262	peptide ABC transporter, ATP-binding protein
SAG1516	255	peptide ABC transporter, permease protein
SAG1517	314	peptide ABC transporter, permease protein
SAG1518	538	peptide ABC transporter, peptide-binding protein
SAG1519	229	ribosomal protein L1
SAG1520	141	ribosomal protein L11
SAG1521	388	transposase, IS30 family, putative
SAG1522	460	transporter, major facilitator family
SAG1523	404	peptidase, M20/M25/M40 family
SAG1524	294	transcriptional regulator, LysR family
SAG1525	117	conserved hypothetical protein
SAG1526	178	IS861, transposase OrfA
SAG1527	277	IS861, transposase OrfB
SAG1528	571	chorismate binding enzyme
SAG1529	816	FtsK/SpoIIIE family protein
SAG1530	267	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
SAG1531	277	manganese ABC transporter, permease protein
SAG1532	238	manganese ABC transporter, ATP-binding protein
SAG1533	308	manganese ABC transporter, manganese-binding adhesion liprotein
SAG1534	215	iron-dependent transcriptional regulator
SAG1535	229	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase
SAG1536	89	conserved hypothetical protein
SAG1537	184	MutT/nudix family protein
SAG1538	459	UDP-N-acetylglucosamine pyrophosphorylase
SAG1539	31	hypothetical protein
SAG1540	137	conserved hypothetical protein
SAG1541	125	glyoxalase family protein
SAG1542	318	oxidoreductase, Gfo/Idh/MocA family
SAG1543	NA	conserved hypothetical protein, authentic frameshift
SAG1544	232	gluconate 5-dehydrogenase, putative
SAG1545	78	conserved hypothetical protein
SAG1546	82	conserved hypothetical protein
SAG1547	166	acetyltransferase, GNAT family
SAG1548	422	glycosyl transferase, group 2 family protein
SAG1549	127	IS1381, transposase OrfA
SAG1550	129	IS1381, transposase OrfB
SAG1551	67	hypothetical protein
SAG1552	719	conserved hypothetical protein
SAG1553	477	hypothetical protein
SAG1554	225	hypothetical protein
SAG1555	231	hypothetical protein
SAG1556	445	branched-chain amino acid transport system II carrier protein

Table 1: Complet list of GBS predicted genes

ORF	Siz (a.a.)	Annotation
SAG1557	665	methionyl-tRNA synthetase
SAG1558	291	tellurite resistance protein TehB
SAG1559	231	membrane protein, putative
SAG1560	40	hypothetical protein
SAG1561	405	PTS system, IIC component, putative
SAG1562	280	conserved hypothetical protein
SAG1563	275	exodeoxyribonuclease
SAG1564	118	conserved hypothetical protein
SAG1565	158	methylated-DNA--protein-cysteine S-methyltransferase
SAG1566	393	D-isomer specific 2-hydroxyacid dehydrogenase family protein
SAG1567	182	acetyltransferase, GNAT family
SAG1568	NA	phosphoserine aminotransferase, authentic frameshift
SAG1569	211	copper homeostasis protein CutC, putative
SAG1570	34	conserved hypothetical protein
SAG1571	53	hypothetical protein
SAG1572	287	tetrapyrrole methylase family protein
SAG1573	108	conserved hypothetical protein
SAG1574	287	DNA polymerase III, delta prime subunit, putative
SAG1575	211	thymidylate kinase
SAG1576	267	transposase, IS30 family, putative, truncation
SAG1577	219	AcuB family protein
SAG1578	236	branched-chain amino acid ABC transporter, ATP-binding protein
SAG1579	254	branched-chain amino acid ABC transporter, ATP-binding protein
SAG1580	317	branched-chain amino acid ABC transporter, permease protein
SAG1581	289	branched-chain amino acid ABC transporter, permease protein
SAG1582	388	branched-chain amino acid ABC transporter, amino acid-binding protein
SAG1583	81	conserved hypothetical protein
SAG1584	377	IS1548, transposase
SAG1585	196	ATP-dependent Clp protease, proteolytic subunit ClpP
SAG1586	209	uracil phosphoribosyltransferase
SAG1587	389	aminotransferase, class I
SAG1588	182	RNA methyltransferase, TrmH family, group 2
SAG1589	450	amino acid permease, putative
SAG1590	449	potassium uptake protein, Trk family
SAG1591	475	cation uptake protein, Trk family
SAG1592	83	conserved hypothetical protein TIGR00278
SAG1593	240	ribosomal large subunit pseudouridine synthase B
SAG1594	194	conserved hypothetical protein TIGR00281
SAG1595	235	conserved hypothetical protein
SAG1596	246	integrase/recombinase, phage integrase family
SAG1597	157	CBS domain protein
SAG1598	173	conserved hypothetical protein
SAG1599	324	HAM1 protein
SAG1600	264	glutamate racemase
SAG1601	79	conserved hypothetical protein
SAG1602	180	membrane protein, putative
SAG1603	173	transcriptional regulator, biotin repressor family

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1604	229	membrane protein, putative
SAG1605	167	conserved hypothetical protein
SAG1606	247	RNA methyltransferase, TrmH family
SAG1607	92	acylphosphatase
SAG1608	310	lipoprotein, putative
SAG1609	221	amino acid ABC transporter, permease protein
SAG1610	285	amino acid ABC transporter, substrate-binding protein
SAG1611	486	amidase family protein
SAG1612	160	transcription elongation factor GreA
SAG1613	600	conserved hypothetical protein
SAG1614	167	acetyltransferase, GNAT family
SAG1615	443	UDP-N-acetylmuramate--alanine ligase
SAG1616	205	conserved hypothetical protein
SAG1617	32	hypothetical protein
SAG1618	1032	Snf2 family protein
SAG1619	377	IS1548, transposase
SAG1620	436	phosphoglycerate dehydrogenase-related protein
SAG1621	300	primosomal protein DnaI
SAG1622	391	conserved hypothetical protein
SAG1623	159	conserved hypothetical protein TIGR00244
SAG1624	501	sensor histidine kinase CsrS
SAG1625	229	DNA-binding response regulator CsrR
SAG1626	177	conserved hypothetical protein
SAG1627	296	heat shock protein HtpX
SAG1628	184	lemA protein
SAG1629	237	glucose-inhibited division protein B
SAG1630	459	sodium transport family protein
SAG1631	223	potassium uptake protein, Trk family, putative
SAG1632	276	cobalt transport family protein
SAG1633	558	ABC transporter, ATP-binding protein
SAG1634	212	conserved hypothetical protein
SAG1635	402	sodium:dicarboxylate symporter family protein
SAG1636	455	branched-chain amino acid transport system II carrier protein
SAG1637	351	alcohol dehydrogenase, zinc-containing
SAG1638	230	ABC transporter, permease protein
SAG1639	356	ABC transporter, ATP-binding protein
SAG1640	458	peptidase, M20/M25/M40 family
SAG1641	274	YaeC family protein
SAG1642	277	ABC transporter, substrate-binding protein
SAG1643	229	glutamine amidotransferase, class I
SAG1644	37	hypothetical protein
SAG1645	238	conserved hypothetical protein TIGR01033
SAG1646	32	hypothetical protein
SAG1647	328	dihydroxyacetone kinase family protein
SAG1648	178	transcriptional regulator, TetR family, putative
SAG1649	37	hypothetical protein
SAG1650	329	dihydroxyacetone kinase family protein
SAG1651	192	dihydroxyacetone kinase family protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1652	124	conserved hypothetical protein
SAG1653	237	glycerol uptake facilitator protein
SAG1654	134	conserved hypothetical protein
SAG1655	237	transcriptional regulator, MerR family
SAG1656	369	conserved hypothetical protein
SAG1657	83	hypothetical protein
SAG1658	244	conserved hypothetical protein
SAG1659	118	iojap-related protein
SAG1660	173	isochorismatase family protein
SAG1661	195	conserved hypothetical protein TIGR00488
SAG1662	210	conserved hypothetical protein TIGR00482
SAG1663	105	conserved hypothetical protein TIGR00253
SAG1664	372	GTP-binding protein
SAG1665	177	hydrolase, haloacid dehalogenase-like family
SAG1666	304	membrane protein, putative
SAG1667	480	glutamyl-tRNA(Gln) amidotransferase, B subunit
SAG1668	488	glutamyl-tRNA(Gln) amidotransferase, A subunit
SAG1669	100	glutamyl-tRNA(Gln) amidotransferase, C subunit
SAG1670	881	pyruvate phosphate dikinase
SAG1671	276	protein of unknown function
SAG1672	170	CBS domain protein
SAG1673	321	3-hydroxyacyl-CoA dehydrogenase family protein
SAG1674	182	isochorismatase family protein
SAG1675	261	transcriptional regulator CodY, putative
SAG1676	403	aminotransferase, class I
SAG1677	150	conserved hypothetical protein
SAG1678	460	hydrolase, haloacid dehalogenase-like family
SAG1679	320	asparaginase family protein
SAG1680	292	shikimate 5-dehydrogenase
SAG1681	304	oxidoreductase, aldo/keto reductase family
SAG1682	671	ATP-dependent DNA helicase RecG
SAG1683	512	immunogenic secreted protein, putative
SAG1684	366	alanine racemase
SAG1685	119	holo-(acyl-carrier-protein) synthase
SAG1686	335	phospho-2-dehydro-3-deoxyheptonate aldolase
SAG1687	842	preprotein translocase, SecA subunit
SAG1688	315	mannose-6-phosphate isomerase, class I
SAG1689	293	fructokinase
SAG1690	639	PTS system, IIABC components
SAG1691	479	sucrose-6-phosphate hydrolase
SAG1692	320	sucrose operon repressor ScrR
SAG1693	144	N utilization substance protein B
SAG1694	129	conserved hypothetical protein
SAG1695	186	translation elongation factor P
SAG1696	38	hypothetical protein
SAG1697	48	hypothetical protein
SAG1698	99	conserved hypothetical protein
SAG1699	30	hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1700	76	hypothetical protein
SAG1701	56	hypothetical protein
SAG1702	41	hypothetical protein
SAG1703	54	hypothetical protein
SAG1704	150	cytidine/deoxycytidylate deaminase family protein
SAG1705	NA	peptidase, M24 family, authentic point mutation
SAG1706	238	conserved hypothetical protein
SAG1707	499	drug resistance transporter, EmrB/QacA family
SAG1708	38	hypothetical protein
SAG1709	942	excinuclease ABC, A subunit
SAG1710	223	conserved hypothetical protein
SAG1711	314	magnesium transporter, CorA family
SAG1712	79	ribosomal protein S18
SAG1713	163	single-strand binding protein
SAG1714	95	ribosomal protein S6
SAG1715	374	A/G-specific adenine glycosylase
SAG1716	197	transcriptional regulator, Cro/CI family
SAG1717	104	thioredoxin
SAG1718	166	PAP2 family protein
SAG1719	779	MutS2 family protein
SAG1720	180	conserved hypothetical protein
SAG1721	103	conserved hypothetical protein
SAG1722	297	ribonuclease HIII
SAG1723	197	signal peptidase I
SAG1724	806	helicase, putative
SAG1725	160	conserved hypothetical protein
SAG1726	364	DNA-damage-inducible protein P
SAG1727	770	formate acetyltransferase
SAG1728	124	FMN-binding protein
SAG1729	309	conserved hypothetical protein
SAG1730	251	conserved hypothetical protein
SAG1731	298	membrane protein, putative
SAG1732	282	glycerol uptake facilitator protein, putative
SAG1733	150	universal stress protein family
SAG1734	400	transporter, putative
SAG1735	219	transcriptional regulator, Crp/Fnr family
SAG1736	761	X-pro dipeptidyl-peptidase
SAG1737	119	hypothetical protein
SAG1738	326	polyprenyl synthetase family protein
SAG1739	582	ABC transporter, ATP-binding protein CydC
SAG1740	572	ABC transporter, ATP-binding protein CydD
SAG1741	339	cytochrome d ubiquinol oxidase, subunit II
SAG1742	475	cytochrome d oxidase, subunit I
SAG1743	402	pyridine nucleotide-disulphide oxidoreductase family protein
SAG1744	299	prenyltransferase, UbiA family
SAG1745	148	hypothetical protein
SAG1746	35	hypothetical protein
SAG1747	99	conserved hypothetical protein TIGR00103

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1748	396	cyclopropane-fatty-acyl-phospholipid synthase
SAG1749	241	transcriptional regulator, merR family
SAG1750	195	exonuclease
SAG1751	178	conserved hypothetical protein
SAG1752	390	conserved hypothetical protein TIGR00275
SAG1753	260	conserved hypothetical protein
SAG1754	89	ribosomal protein S14
SAG1755	38	hypothetical protein
SAG1756	341	conserved hypothetical protein
SAG1757	336	O-sialoglycoprotein endopeptidase family protein
SAG1758	135	ribosomal-protein-alanine acetyltransferase, putative
SAG1759	230	protein of unknown function
SAG1760	76	conserved hypothetical protein
SAG1761	559	metallo-beta-lactamase superfamily protein
SAG1762	169	conserved hypothetical protein
SAG1763	448	glutamine synthetase, type I
SAG1764	123	transcriptional regulator GlnR
SAG1765	179	conserved hypothetical protein
SAG1766	398	phosphoglycerate kinase
SAG1767	289	acid phosphatase
SAG1768	336	glyceraldehyde 3-phosphate dehydrogenase
SAG1769	692	translation elongation factor G
SAG1770	156	ribosomal protein S7
SAG1771	137	ribosomal protein S12
SAG1772	270	pur operon repressor
SAG1773	313	HD domain protein
SAG1774	424	conserved hypothetical protein
SAG1775	210	conserved hypothetical protein
SAG1776	220	ribulose-phosphate 3-epimerase
SAG1777	290	conserved hypothetical protein TIGR00157
SAG1778	283	rRNA (guanine-N1-)-methyltransferase, putative
SAG1779	290	dimethyladenosine transferase
SAG1780	163	hypothetical protein
SAG1781	186	primase-related protein
SAG1782	260	deoxyribonuclease, TatD family
SAG1783	90	hypothetical protein
SAG1784	130	hypothetical protein
SAG1785	430	hypothetical protein
SAG1786	130	protein of unknown function
SAG1787	420	dltD protein
SAG1788	79	D-alanyl carrier protein
SAG1789	421	dltB protein
SAG1790	511	D-alanine-activating enzyme
SAG1791	395	sensor histidine kinase
SAG1792	224	DNA-binding response regulator
SAG1793	44	ribosomal protein L34
SAG1794	451	membrane protein, putative
SAG1795	388	transposase, IS30 family, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1796	575	amino acid ABC transporter, permease protein
SAG1797	407	amino acid ABC transporter, ATP-binding protein
SAG1798	39	hypothetical protein
SAG1799	792	xylulose-5-phosphate/fructose-6-phosphate phosphoketolase
SAG1800	363	conserved hypothetical protein
SAG1801	559	transcriptional antiterminator, BglG family
SAG1802	253	conserved hypothetical protein
SAG1803	505	carbohydrate kinase, FGGY family
SAG1804	329	hypothetical protein
SAG1805	483	PTS system, IIC component, putative
SAG1806	318	glyoxylate reductase, NADH-dependent
SAG1807	339	hypothetical protein
SAG1808	327	sugar binding transcriptional regulator, LacI family
SAG1809	215	transaldolase family protein
SAG1810	238	carbohydrate isomerase, AraD/FucA family
SAG1811	287	hexulose-6-phosphate isomerase, putative
SAG1812	221	hexulose-6-phosphate synthase, putative
SAG1813	161	PTS system, IIA component
SAG1814	92	PTS system, IIB component
SAG1815	479	transport protein SgaT, putative
SAG1816	205	hypothetical protein
SAG1817	157	hypothetical protein
SAG1818	430	adenylosuccinate synthetase
SAG1819	340	perfringolysin O regulator protein
SAG1820	224	conserved hypothetical protein
SAG1821	750	glutamate--cysteine ligase/amino acid ligase, putative
SAG1822	272	protein of unknown function
SAG1823	418	protein of unknown function
SAG1824	291	chaperonin, 33 kDa
SAG1825	325	NifR3/Smm1 family protein
SAG1826	213	deoxynucleoside kinase family protein
SAG1827	163	phosphinothricin N-acetyltransferase
SAG1828	815	ATP-dependent Clp protease, ATP-binding subunit
SAG1829	154	transcriptional regulator CtsR
SAG1830	153	conserved hypothetical protein
SAG1831	346	translation elongation factor Ts
SAG1832	256	ribosomal protein S2
SAG1833	186	alkyl hydroperoxide reductase, subunit C
SAG1834	510	alkyl hydroperoxide reductase, subunit F
SAG1835	134	conserved hypothetical protein
SAG1836	61	conserved hypothetical protein
SAG1837	468	prophage LambdaSa2, lysin, putative
SAG1838	109	prophage LambdaSa2, holin, putative
SAG1839	136	conserved hypothetical protein
SAG1840	112	hypothetical protein
SAG1841	76	conserved domain protein
SAG1842	1224	prophage LambdaSa2, PblB, putative
SAG1843	240	conserved hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1844	911	conserved hypothetical protein
SAG1845	42	hypothetical protein
SAG1846	158	hypothetical protein
SAG1847	227	conserved hypothetical protein
SAG1848	114	conserved hypothetical protein
SAG1849	115	hypothetical protein
SAG1850	101	hypothetical protein
SAG1851	111	conserved domain protein
SAG1852	420	conserved domain protein
SAG1853	180	prophage LambdaSa2, protease, putative
SAG1854	380	conserved hypothetical protein
SAG1855	570	prophage LambdaSa2, terminase large subunit, putative
SAG1856	161	hypothetical protein
SAG1857	119	prophage LambdaSa2, HNH endonuclease family protein
SAG1858	95	hypothetical protein
SAG1859	180	prophage LambdaSa2, site-specific recombinase, phage integrase family
SAG1860	154	conserved hypothetical protein
SAG1861	119	prophage LambdaSa2, transcriptional regulator, Cro/CI family
SAG1862	86	hypothetical protein
SAG1863	138	prophage LambdaSa2, single-strand binding protein
SAG1864	68	hypothetical protein
SAG1865	74	conserved hypothetical protein
SAG1866	109	conserved hypothetical protein
SAG1867	163	conserved hypothetical protein
SAG1868	134	hypothetical protein
SAG1869	437	prophage LambdaSa2, type II DNA modification methyltransferase, putative
SAG1870	273	prophage LambdaSa2, DNA replication protein DnaC, putative
SAG1871	248	prophage LambdaSa2, bacteriophage replication protein/hypothetical protein, truncation/fusion
SAG1872	200	hypothetical protein
SAG1873	443	prophage LambdaSa2, replicative DNA helicase
SAG1874	87	hypothetical protein
SAG1875	94	conserved hypothetical protein
SAG1876	176	prophage LambdaSa2, HNH endonuclease family protein
SAG1877	236	prophage LambdaSa2, antirepressor protein, putative
SAG1878	102	conserved domain protein
SAG1879	156	hypothetical protein
SAG1880	54	hypothetical protein
SAG1881	51	hypothetical protein
SAG1882	120	prophage LambdaSa2, repressor protein, putative
SAG1883	128	conserved hypothetical protein
SAG1884	134	hypothetical protein
SAG1885	356	prophage LambdaSa2, site-specific recombinase, phage integrase family
SAG1886	32	hypothetical protein
SAG1887	689	Na ⁺ /H ⁺ exchanger family protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1888	78	hypothetical protein
SAG1889	317	microcin immunity protein MccF, putative
SAG1890	631	endopeptidase O
SAG1891	327	oxidoreductase, Gfo/Idh/MocA family
SAG1892	358	membrane protein, putative
SAG1893	59	hypothetical protein
SAG1894	214	cyclic nucleotide-binding domain protein
SAG1895	204	polypeptide deformylase
SAG1896	333	sugar binding transcriptional regulator RegR
SAG1897	634	conserved hypothetical protein
SAG1898	271	PTS system, IID component
SAG1899	288	PTS system, IIC component
SAG1900	164	PTS system, IIB component
SAG1901	398	glucuronyl hydrolase
SAG1902	144	PTS system, IIA component
SAG1903	34	hypothetical protein
SAG1904	270	oxidoreductase, short-chain dehydrogenase/reductase family
SAG1905	212	conserved hypothetical protein
SAG1906	335	carbohydrate kinase, PfkB family
SAG1907	212	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase
SAG1908	499	hypothetical protein
SAG1909	204	nitroreductase family protein
SAG1910	141	transcriptional regulator, MarR family
SAG1911	1468	DNA polymerase III, alpha subunit, Gram-positive type
SAG1912	194	N-acetylmuramoyl-L-alanine amidase, family 4 protein
SAG1913	617	prolyl-tRNA synthetase
SAG1914	419	membrane-associated zinc metalloprotease, putative
SAG1915	264	phosphatidate cytidyltransferase
SAG1916	250	undecaprenyl diphosphate synthase
SAG1917	113	preprotein translocase, YajC subunit
SAG1918	114	bacteriocin transport accessory protein, putative
SAG1919	387	malate oxidoreductase
SAG1920	445	citrate carrier protein, CCS family
SAG1921	508	sensor histidine kinase
SAG1922	229	response regulator
SAG1923	331	UDP-glucose 4-epimerase
SAG1924	535	glucan 1,6-alpha-glucosidase
SAG1925	377	sugar ABC transporter, ATP-binding protein
SAG1926	283	helix-turn-helix domain protein, fis-type
SAG1927	298	lacX protein
SAG1928	325	tagatose 1,6-diphosphate aldolase
SAG1929	310	tagatose-6-phosphate kinase
SAG1930	171	galactose-6-phosphate isomerase, LacB subunit
SAG1931	141	galactose-6-phosphate isomerase, LacA subunit
SAG1932	816	neuraminidase-related protein
SAG1933	482	PTS system, IIC component, putative
SAG1934	101	PTS system, IIB component, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1935	157	PTS system, IIA component, putative
SAG1936	258	lactose phosphotransferase system repressor
SAG1937	NA	streptococcal histidine triad family protein, degenerate
SAG1938	307	adhesion lipoprotein
SAG1939	147	protein of unknown function TIGR00256
SAG1940	738	GTP pyrophosphokinase family protein
SAG1941	800	2',3'-cyclic-nucleotide 2'-phosphodiesterase
SAG1942	151	nrdI protein
SAG1943	345	conserved hypothetical protein
SAG1944	165	conserved hypothetical protein
SAG1945	345	iron ABC transporter, iron-binding protein
SAG1946	257	DNA-binding response regulator
SAG1947	549	conserved hypothetical protein
SAG1948	275	PTS system, IID component
SAG1949	269	PTS system, IIC component
SAG1950	163	PTS system, IIB component
SAG1951	141	PTS system, IIA component, putative
SAG1952	353	membrane protein, putative
SAG1953	60	hypothetical protein
SAG1954	384	membrane protein, putative
SAG1955	282	ABC transporter, ATP-binding protein
SAG1956	96	conserved hypothetical protein, truncation
SAG1957	250	response regulator
SAG1958	276	conserved hypothetical protein
SAG1959	727	PTS system, IIABC components
SAG1960	551	sensor histidine kinase
SAG1961	225	phosphate regulon response regulator PhoB
SAG1962	218	phosphate transport system regulatory protein PhoU, putative
SAG1963	253	phosphate ABC transporter, ATP-binding protein
SAG1964	292	phosphate ABC transporter, permease protein
SAG1965	281	phosphate ABC transporter, permease protein
SAG1966	293	hemolysin precursor, putative
SAG1967	195	hypothetical protein
SAG1968	246	conserved hypothetical protein TIGR00046
SAG1969	317	ribosomal protein L11 methyltransferase
SAG1970	102	conserved hypothetical protein
SAG1971	41	hypothetical protein
SAG1972	238	transcriptional regulator, MerR family
SAG1973	156	acetyltransferase, GNAT family
SAG1974	152	MutT/nudix family protein
SAG1975	47	hypothetical protein
SAG1976	156	conserved hypothetical protein
SAG1977	163	acetyltransferase, GNAT family
SAG1978	422	ATPase, AAA family
SAG1979	253	membrane protein, putative
SAG1980	300	ABC transporter, ATP-binding protein
SAG1981	68	hypothetical protein
SAG1982	359	transcriptional regulator, Cro/CI family

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG1983	105	conserved hypothetical protein
SAG1984	188	conserved hypothetical protein TIGR00730
SAG1985	51	hypothetical protein
SAG1986	375	site-specific recombinase, phage integrase family
SAG1987	61	conserved hypothetical protein
SAG1988	342	conserved hypothetical protein
SAG1989	139	hypothetical protein
SAG1990	127	hypothetical protein
SAG1991	204	transcriptional regulator, Cro/CI family
SAG1992	518	protein of unknown function
SAG1993	373	site-specific recombinase, phage integrase family
SAG1994	108	conserved hypothetical protein
SAG1995	210	hypothetical protein
SAG1996	263	cell wall surface anchor family protein, putative
SAG1997	182	hypothetical protein
SAG1998	457	hypothetical protein
SAG1999	47	hypothetical protein
SAG2000	666	membrane protein, putative
SAG2001	756	conjugal transfer protein, interruption-C
SAG2002	129	IS1381, transposase OrfB
SAG2003	127	IS1381, transposase OrfA
SAG2004	67	conjugal transfer protein, interruption-N
SAG2005	136	conserved hypothetical protein
SAG2006	88	conserved hypothetical protein
SAG2007	317	conserved hypothetical protein
SAG2008	84	conserved hypothetical protein
SAG2009	88	conserved hypothetical protein
SAG2010	157	hypothetical protein
SAG2011	160	conserved hypothetical protein
SAG2012	90	hypothetical protein
SAG2013	189	hypothetical protein
SAG2014	449	hypothetical protein
SAG2015	99	transcriptional regulator, Cro/CI family
SAG2016	125	hypothetical protein
SAG2017	429	transcriptional regulator, Cro/CI family
SAG2018	553	FtsK/SpoIIIE family protein
SAG2019	153	hypothetical protein
SAG2020	98	hypothetical protein
SAG2021	826	cell wall surface anchor family protein
SAG2022	417	transposase, ISL3 family
SAG2023	546	mercuric reductase
SAG2024	130	mercuric resistance operon regulatory protein MerR
SAG2025	522	Mn2+/Fe2+ transporter, NRAMF family
SAG2026	240	membrane protein, putative
SAG2027	205	ABC transporter, ATP-binding protein
SAG2028	36	conserved hypothetical protein
SAG2029	284	streptomycin resistance protein
SAG2030	130	hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG2031	202	hypothetical protein
SAG2032	111	conserved hypothetical protein
SAG2033	162	acetyltransferase, GNAT family
SAG2034	247	membrane protein, putative
SAG2035	300	ABC transporter, ATP-binding protein
SAG2036	68	hypothetical protein
SAG2037	358	transcriptional regulator, Cro/CI family
SAG2038	204	PAP2 family protein
SAG2039	98	conserved hypothetical protein
SAG2040	186	conserved hypothetical protein TIGR00730
SAG2041	287	protease, putative
SAG2042	100	rhodanese family protein
SAG2043	255	cAMP factor
SAG2044	62	hypothetical protein
SAG2045	179	DNA topology modulation protein FlaR, putative
SAG2046	361	glycerol dehydrogenase, putative
SAG2047	235	conserved hypothetical protein
SAG2048	614	5-methyltetrahydrofolate--homocysteine methyltransferase, putative
SAG2049	745	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
SAG2050	107	conserved hypothetical protein
SAG2051	230	branched-chain amino acid transport protein AzlC, putative
SAG2052	41	hypothetical protein
SAG2053	1570	serine protease, subtilase family, putative
SAG2054	228	DNA-binding response regulator
SAG2055	462	sensor histidine kinase
SAG2056	202	chromosome assembly-related protein
SAG2057	833	leucyl-tRNA synthetase
SAG2058	415	major facilitator family protein
SAG2059	281	protein of unknown function
SAG2060	398	glycosyl transferase, family 8
SAG2061	401	glycosyl transferase, family 8
SAG2062	179	transcription antitermination protein NusG
SAG2063	630	pathogenicity protein, putative
SAG2064	57	preprotein translocase, SecE subunit, putative
SAG2065	50	ribosomal protein L33
SAG2066	773	penicillin-binding protein 2A
SAG2067	294	ribosomal large subunit pseudouridine synthase, RluD subfamily
SAG2068	546	conserved hypothetical protein
SAG2069	403	phosphopentomutase
SAG2070	223	deoxyribose-phosphate aldolase
SAG2071	400	Na ⁺ dependent nucleoside transporter
SAG2072	259	uridine phosphorylase
SAG2073	245	transcriptional regulator, GntR family
SAG2074	540	60 kda chaperonin
SAG2075	94	chaperonin, 10 kDa
SAG2076	267	ABC transporter, ATP-binding protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG2077	298	ABC transporter, permease protein
SAG2078	320	protein of unknown function/lipoprotein, putative
SAG2079	265	hydrolase, haloacid dehalogenase-like family
SAG2080	286	glyoxalase family protein
SAG2081	243	conserved hypothetical protein
SAG2082	205	anaerobic ribonucleoside-triphosphate reductase activating protein
SAG2083	163	acetyltransferase, GNAT family
SAG2084	310	virulence factor MviM, putative
SAG2085	47	conserved hypothetical protein
SAG2086	723	anaerobic ribonucleoside-triphosphate reductase
SAG2087	495	membrane protein, putative
SAG2088	40	hypothetical protein
SAG2089	105	conserved hypothetical protein
SAG2090	136	conserved hypothetical protein TIGR00250
SAG2091	88	conserved hypothetical protein
SAG2092	132	conserved hypothetical protein
SAG2093	379	recA protein
SAG2094	NA	competence/damage-inducible protein CinA, authentic frameshift
SAG2095	183	DNA-3-methyladenine glycosylase I
SAG2096	196	Holliday junction DNA helicase RuvA
SAG2097	418	transporter, putative
SAG2098	659	DNA mismatch repair protein HexB
SAG2099	33	hypothetical protein
SAG2100	67	cold shock protein, CSD family
SAG2101	858	DNA mismatch repair protein HexA
SAG2102	145	arginine repressor ArgR, putative
SAG2103	563	arginyl-tRNA synthetase
SAG2104	102	conserved hypothetical protein
SAG2105	290	conserved hypothetical protein
SAG2106	314	conserved hypothetical protein
SAG2107	583	aspartyl-tRNA synthetase
SAG2108	426	histidyl-tRNA synthetase
SAG2109	60	ribosomal protein L32
SAG2110	49	ribosomal protein L33
SAG2111	173	conserved hypothetical protein
SAG2112	494	site-specific recombinase, phage integrase family
SAG2113	82	conserved hypothetical protein
SAG2114	342	conserved hypothetical protein
SAG2115	143	hypothetical protein
SAG2116	151	conserved hypothetical protein
SAG2117	71	hypothetical protein
SAG2118	306	transcriptional regulator, Cro/CI family
SAG2119	373	conserved domain protein
SAG2120	269	hypothetical protein
SAG2121	223	hypothetical protein
SAG2122	223	DNA-binding response regulator
SAG2123	454	sensor histidine kinase
SAG2124	517	membrane protein, putative

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG2125	308	carbamate kinase
SAG2126	332	ornithine carbamoyltransferase
SAG2127	431	sensor histidine kinase
SAG2128	277	response regulator
SAG2129	240	amino acid ABC transporter, ATP-binding protein
SAG2130	504	amino acid ABC transporter, amino acid-binding protein/permease protein
SAG2131	847	membrane protein, putative
SAG2132	247	conserved hypothetical protein
SAG2133	118	conserved hypothetical protein
SAG2134	772	membrane protein, putative
SAG2135	179	transcriptional regulator, TetR family, putative
SAG2136	98	conserved hypothetical protein
SAG2137	203	ribosomal protein S4
SAG2138	95	conserved hypothetical protein
SAG2139	451	replicative DNA helicase
SAG2140	150	ribosomal protein L9
SAG2141	660	DHH family protein
SAG2142	613	glucose inhibited division protein A
SAG2143	203	membrane protein, putative
SAG2144	373	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
SAG2145	222	L-serine dehydratase, iron-sulfur-dependent, beta subunit
SAG2146	290	L-serine dehydratase, iron-sulfur-dependent, alpha subunit
SAG2147	234	protein of unknown function/lipoprotein, putative
SAG2148	179	LysM domain protein
SAG2149	264	cobalt transport family protein
SAG2150	280	ABC transporter, ATP-binding protein
SAG2151	279	ABC transporter, ATP-binding protein
SAG2152	180	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
SAG2153	427	peptidase, M16 family
SAG2154	414	conserved hypothetical protein
SAG2155	117	conserved hypothetical protein
SAG2156	369	recF protein
SAG2157	278	transporter, putative
SAG2158	220	transcriptional regulator, Cro/CI family
SAG2159	493	inosine-5'-monophosphate dehydrogenase
SAG2160	161	transcriptional regulator, ArgR family
SAG2161	226	transcriptional regulator, Crp/Fnr family
SAG2162	234	conserved hypothetical protein
SAG2163	410	arginine deiminase
SAG2164	136	acetyltransferase, GNAT family
SAG2165	337	ornithine carbamoyltransferase
SAG2166	475	arginine/ornithine antiporter
SAG2167	318	carbamate kinase
SAG2168	341	tryptophanyl-tRNA synthetase
SAG2169	230	membrane protein, putative
SAG2170	290	conserved hypothetical protein

Table 1: Complete list of GBS predicted genes

ORF	Size (a.a.)	Annotation
SAG2171	539	ABC transporter, ATP-binding protein
SAG2172	859	ABC transporter, permease protein, putative
SAG2173	159	conserved hypothetical protein TIGR00246
SAG2174	409	serine protease
SAG2175	257	partitioning protein, ParB family

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0017	447	+							pcsB
SAG0031	299	+							peptidase, M23/M37 family
SAG0032	434	+				+	+		group B streptococcal surface immunogenic protein
SAG0034	438	+		+		+	+		sugar ABC transporter, sugar-binding protein
SAG0051	126	+				+	+		MORN motif family protein
SAG0079	212				+	+	+		adenylate kinase
SAG0086	85			+				+	lipoprotein, putative
SAG0093	250	+				+	+		D-alanyl-D-alanine carboxypeptidase family protein
SAG0094	191	+							N-acetylmuramoyl-L-alanine amidase, family 4 protein
SAG0108	308	+							conserved hypothetical protein
SAG0114	322	+		+					ribose ABC transporter, periplasmic D-ribose-binding protein
SAG0124	356	+							sensor histidine kinase
SAG0132	294	+				+	+		SPFH domain/Band 7 family protein
SAG0134	96	+						+	hypothetical protein
SAG0146	395	+							penicillin-binding protein 4, putative
SAG0147	411	+							D-alanyl-D-alanine carboxypeptidase family protein
SAG0148	551			+		+	-		oligopeptide ABC transporter, substrate-binding protein, putative
SAG0166	123	+							conserved domain protein
SAG0176	94	+							conserved hypothetical protein
SAG0187	542	+		+		+	+		oligopeptide ABC transporter, oligopeptide-binding protein
SAG0206	60			+				+	lipoprotein, putative
SAG0213	39	+						+	hypothetical protein
SAG0231	135	+							hypothetical protein
SAG0242	308			+		+	-		amino acid ABC transporter, amino acid-binding protein
SAG0245	152			+		+	-	+	protein of unknown function/lipoprotein, putative
SAG0255	315	+							conserved hypothetical protein
SAG0257	53			+				+	lipoprotein, putative
SAG0265	235	+				+	-	+	conserved hypothetical protein
SAG0290	270	+				+	+		ABC transporter, substrate-binding protein
SAG0298	750	+							penicillin-binding protein 1A

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0306	535	+							KH domain protein
SAG0321	339	+							sensor histidine kinase, putative
SAG0329	106	+							PTS system, cellobiose-specific IIB component
SAG0368	435	+				+	+		protein of unknown function
SAG0371	167	+						+	hypothetical protein
SAG0383	334	+		+		+	-		protein of unknown function/lipoprotein, putative
SAG0392	521	+	+			+	+		cell wall surface anchor family protein
SAG0394	345				+				sensor histidine kinase
SAG0405	347	+		+		+	+		protein of unknown function/lipoprotein, putative
SAG0406	299	+							UTP-glucose-1-phosphate uridylyltransferase
SAG0407	338	+							glycerol-3-phosphate dehydrogenase (NAD(P)+)
SAG0416	1233	+	+			+	+		protease, putative
SAG0421	1055		+			+	-		cell wall surface anchor family protein
SAG0433	1389		+						surface protein Rib
SAG0437	123			+					lipoprotein, putative
SAG0451	149	+		+				+	bacteriocin transport accessory protein, putative
SAG0455	357	+							conserved hypothetical protein
SAG0472	126	+				+	-		rhodanese-like family protein
SAG0482	84	+							YGGT family protein
SAG0499	275				+				hemolysin A
SAG0503	279	+				+	+		lipase/acylhydrolase
SAG0504	200	+							conserved hypothetical protein
SAG0506	65	+						+	hypothetical protein
SAG0521	236	+							carboxymethylenebutenolidase-related protein
SAG0535	506	+				+	+		zinc ABC transporter, zinc-binding adhesion liprotein
SAG0596	670				+				prophage LambdaSal, pblA protein, internal deletion
SAG0603	111				+				conserved hypothetical protein
SAG0604	239				+				prophage LambdaSal, lysin, putative
SAG0617	439				+				sensor histidine kinase VncS
SAG0624	574	+							septation ring formation regulator EzrA, putative
SAG0629	354	+							conserved domain protein
SAG0635	245	+				+	-		acid phosphatase, class B
SAG0638	109	+							cell wall surface anchor family protein, interruption-N

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0645	554		+			+	+		cell wall surface anchor family protein
SAG0646	307	+	+			+	-		cell wall surface anchor family protein
SAG0647	305	+							sortase family protein
SAG0649	890		+			+	+		cell wall surface anchor family protein, putative
SAG0658	383	+		+					lipoprotein, putative
SAG0675	171	+							putative secreted protein
SAG0676	885				+				proteinase, putative
SAG0677	1062		+						hypothetical protein
SAG0679	343	+		+		+	-		protein of unknown function
SAG0680	339	+				+	-		protein of unknown function
SAG0681	353	+							conserved domain protein
SAG0686	261	+				+	+		DNA-entry nuclease, putative
SAG0714	188	+						+	conserved hypothetical protein
SAG0717	266	+				+	+		amino acid ABC transporter, amino acid-binding protein
SAG0720	449				+				sensory box histidine kinase
SAG0738	132	+							conserved hypothetical protein
SAG0739	143	+							conserved hypothetical protein
SAG0742	428				+	+	+		peptidase, U32 family
SAG0755	282	+							peptidase, U32 family
SAG0757	129	+		+		+	-		protein of unknown function/lipoprotein, putative
SAG0764	230				+	+	+		phosphoglycerate mutase family protein
SAG0765	681	+							penicillin-binding protein 2b
SAG0771	512	+	+			+	+	+	cell wall surface anchor family protein
SAG0776	276	+		+					YacC family protein, putative
SAG0777	528				+	+	+		ATP-dependent RNA helicase, DEAD/DEAH box family
SAG0785	330	+							conserved hypothetical protein
SAG0808	309	+		+		+	+		protease maturation protein, putative
SAG0824	417	+							polysaccharide deacetylase family protein
SAG0832	753	+				+	+		protein of unknown function
SAG0833	181	+						+	hypothetical protein
SAG0867	63	+							conserved hypothetical protein
SAG0868	285	+				+	-		DNA-entry nuclease
SAG0886	319	+				+	+		protein of unknown function

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0904	56	+						+	hypothetical protein
SAG0907	877	+		+		+	-		protein of unknown function/lipoprotein, putative
SAG0926	333	+							Tn916, NLP/P60 family protein
SAG0942	185	+				+	+		signal peptidase I, putative
SAG0949	276	+		+		+	+		amino acid ABC transporter, amino acid-binding protein
SAG0954	349			+		+	-		protein of unknown function/lipoprotein, putative
SAG0961	247	+				+	-		sortase SrtA
SAG0963	320	+							conserved hypothetical protein
SAG0971	282	+		+		+	-		protein of unknown function/lipoprotein, putative
SAG0973	320	+						+	nisin-resistance protein, putative
SAG0977	312				+				sensor histidine kinase
SAG0979	553	+		+		+	-		ABC transporter, substrate-binding protein
SAG0984	437	+							sensor histidine kinase CraH
SAG0992	286	+		+		+	+		phosphate ABC transporter, phosphate-binding protein
SAG1007	342	+		+		+	-		iron-compound ABC transporter, iron-compound-binding protein
SAG1014	190	+				-	-		conserved hypothetical protein
SAG1018	40			+				+	lipoprotein, putative
SAG1024	183	+		+					lipoprotein, putative
SAG1029	101	+							hypothetical protein
SAG1030	304	+				+	+		protein of unknown function
SAG1037	157	+						+	hypothetical protein
SAG1052	47		+					+	cell wall surface anchor family protein, putative
SAG1072	200	+							conserved hypothetical protein
SAG1094	278				+	+	+		conserved hypothetical protein
SAG1108	357	+				+	-		spermidine/putrescine ABC transporter, spermidine/putrescine-binding prot.
SAG1121	295	+							polysaccharide deacetylase family protein
SAG1126	228	+				+	+		protein of unknown function
SAG1127	446	+						+	conserved domain protein
SAG1130	49	+						+	hypothetical protein
SAG1138	64	+							conserved hypothetical protein
SAG1139	193	+							conserved hypothetical protein

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG1149	207	+		+					lipoprotein, putative
SAG1184	236	+							conserved hypothetical protein
SAG1186	553				+				metallo-beta-lactamase superfamily protein
SAG1189	334	+							conserved hypothetical protein
SAG1190	551				+				adherence and virulence protein A
SAG1197	1072	+							hyaluronidase
SAG1201	367	+							aminodiacetate oxidase, putative
SAG1206	854	+							conserved domain protein
SAG1214	58	+							hypothetical protein
SAG1216	1252		+			+	-		pullulanase, putative
SAG1227	198	+				+	-		protein of unknown function
SAG1233	822	+				+	-		streptococcal histidine triad family protein
SAG1234	306	+		+		+	+		laminin-binding surface protein
SAG1238	202	+							hypothetical protein
SAG1283	1631		+			+	+		agglutinin receptor
SAG1313	56	+							conserved hypothetical protein
SAG1327	409	+							sensor histidine kinase
SAG1331	979	+	+			+	+		R5 protein
SAG1333	690	+	+			+	+		5'-nucleotidase family protein
SAG1350	544	+							surface antigen-related protein
SAG1361	414	+							conserved hypothetical protein
SAG1371	392	+							conserved hypothetical protein
SAG1393	310			+					iron compound ABC transporter, substrate-binding protein
SAG1404	308	+	+			+	-		cell wall surface anchor family protein
SAG1405	294	+			+	+	+		sortase family protein
SAG1406	293	+							sortase family protein
SAG1407	705	+	+			+	+		cell wall surface anchor family protein
SAG1408	901		+						cell wall surface anchor family protein
SAG1419	577			+				+	lipoprotein, putative
SAG1431	268			+					amino acid ABC transporter, amino acid-binding protein
SAG1433	375	+							conserved hypothetical protein
SAG1441	415	+				+	+		maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG1462	970		+						cell wall surface anchor family protein
SAG1473	192	+	+					+	cell wall surface anchor family protein
SAG1474	680	+	+						amidase family protein
SAG1483	78	+							preprotein translocase, SecY subunit
SAG1488	195	+				+	+		dephospho-CoA kinase
SAG1491	530	+						+	hypothetical protein
SAG1508	590				+	+	-		67 kDa Myosin-crossreactive streptococcal antigen
SAG1518	538	+		+					peptide ABC transporter, peptide-binding protein
SAG1530	267	+		+		+	-		peptidyl-prolyl cis-trans isomerase, cyclophilin-type
SAG1533	308	+		+		+	-		manganese ABC transporter, manganese-binding adhesion liprotein
SAG1544	232	+							gluconate 5-dehydrogenase, putative
SAG1551	67	+						+	hypothetical protein
SAG1552	719	+							conserved hypothetical protein
SAG1553	477	+						+	hypothetical protein
SAG1562	280	+							conserved hypothetical protein
SAG1582	388	+		+		+	-		branched-chain amino acid ABC transporter, amino acid- binding protein
SAG1590	449				+	+	+		potassium uptake protein, Trk family
SAG1601	79	+							conserved hypothetical protein
SAG1610	285			+		+	-		amino acid ABC transporter, substrate-binding protein
SAG1618	1032				+	+	+		Snf2 family protein
SAG1624	501	+							sensor histidine kinase CsrS
SAG1628	184	+							lemA protein
SAG1631	223	+				+	-		potassium uptake protein, Trk family, putative
SAG1641	274	+				+	-		YaeC family protein
SAG1642	277	+		+		+	-		ABC transporter, substrate-binding protein
SAG1683	512	+							immunogenic secreted protein, putative
SAG1706	238	+							conserved hypothetical protein
SAG1745	148	+						+	hypothetical protein
SAG1752	390	+							conserved hypothetical protein TIGR00275
SAG1759	230				+	+	+		protein of unknown function
SAG1762	169	+							conserved hypothetical protein

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG1767	289	+		+					acid phosphatase
SAG1768	336				+	+	+		glyceraldehyde 3-phosphate dehydrogenase
SAG1774	424	+							conserved hypothetical protein
SAG1786	130	+				+	-		protein of unknown function
SAG1787	420	+							ditD protein
SAG1791	395	+							sensor histidine kinase
SAG1822	272	+				+	-		protein of unknown function
SAG1823	418				+	+	+		protein of unknown function
SAG1837	468				+				prophage LambdaSa2, lysin, putative
SAG1838	109	+							prophage LambdaSa2, holin, putative
SAG1839	136	+							conserved hypothetical protein
SAG1842	1224				+				prophage LambdaSa2, PblB, putative
SAG1912	194	+							N-acetylmuramoyl-L-alanine amidase, family 4 protein
SAG1921	508	+							sensor histidine kinase
SAG1932	816	+							neuraminidase-related protein
SAG1938	307	+		+		+	-		adhesion lipoprotein
SAG1941	800	+	+			+	-		2',3'-cyclic-nucleotide 2'-phosphodiesterase
SAG1945	345	+							iron ABC transporter, iron-binding protein
SAG1947	549				+				conserved hypothetical protein
SAG1960	551				+	+	+		sensor histidine kinase
SAG1966	293			+		+	-		hemolysin precursor, putative
SAG1996	263	+	+						cell wall surface anchor family protein, putative
SAG1997	182	+							hypothetical protein
SAG1998	457	+							hypothetical protein
SAG2021	826		+						cell wall surface anchor family protein
SAG2043	255	+							cAMP factor
SAG2053	1570	+	+						serine protease, subtilase family, putative
SAG2055	462				+				sensor histidine kinase
SAG2056	202	+						+	chromosome assembly-related protein
SAG2063	630	+	+						pathogenicity protein, putative
SAG2078	320	+		+		+	-		protein of unknown function/lipoprotein, putative
SAG2094		+				+	+		competence/damage-inducible protein ClnA, authentic frameshift

Table 2

ORF	Size (aa)	Signal Peptide	Sortase motif	Lipo- protein	Other	Western blot	FACS	GBS specific	Annotation
SAG2121	223	+						+	hypothetical protein
SAG2123	454	+							sensor histidine kinase
SAG2141	660	+				+	-		DHH family protein
SAG2147	234	+		+		+	+		protein of unknown function/lipoprotein, putative
SAG2148	179	+							LysM domain protein
SAG2174	409	+							serine protease
SAG0013	428	+				+	-		protein of unknown function

Table 3

ORF	Annotation
SAG0038	conserved hypothetical protein
SAG0048	transcriptional regulator Cro/CI family
SAG0091	transcriptional regulator ComX1 putative
SAG0137	conserved hypothetical protein
SAG0686	DNA-entry nuclease putative
SAG0770	membrane protein putative
SAG0868	DNA-entry nuclease
SAG1143	conserved hypothetical protein
SAG1233	streptococcal histidine triad family protein
SAG1596	integrase/recombinase phage integrase family
SAG1616	conserved hypothetical protein
SAG1721	conserved hypothetical protein.

Table 4: Probable recently duplicated genes

Probable recently duplicated genes are indicated on the same line and are separated by a semicolon.

SAG0148 oligopeptide ABC transporter, substrate-binding protein, putative; SAG0979 ABC transporter, substrate-binding protein
SAG0151 oligopeptide ABC transporter, ATP-binding protein; SAG1515 peptide ABC transporter, ATP-binding protein
SAG0195 IS1548, transposase; SAG0693 IS1548, transposase; SAG0760 IS1548, transposase; SAG0945 IS1548, transposase; SAG1584 IS1548, transposase; SAG1619 IS1548, transposase
SAG0230 conserved hypothetical protein; SAG1039 conserved hypothetical protein
SAG0233 hypothetical protein; SAG1785 hypothetical protein
SAG0261 IS1381, transposase OrfB; SAG0542 IS1381, transposase OrfA; SAG0543 IS1381, transposase OrfB; SAG0966 IS1381, transposase OrfB; SAG1457 IS1381, transposase OrfB; SAG1550 IS1381, transposase OrfB; SAG2002 IS1381, transposase OrfB
SAG0262 IS1381, transposase OrfA; SAG0965 IS1381, transposase OrfA; SAG1549 IS1381, transposase OrfA; SAG1458 IS1381, transposase OrfA; SAG2003 IS1381, transposase OrfA
SAG0383 protein of unknown function/lipoprotein, putative; SAG0785 conserved hypothetical protein
SAG0405 protein of unknown function/lipoprotein, putative; SAG0954 protein of unknown function/lipoprotein, putative
SAG0417 glycosyl transferase, group 2 family protein; SAG1422 glycosyl transferase, group 2 family protein
SAG0429 oxidoreductase, aldo/keto reductase family; SAG1476 oxidoreductase, aldo/keto reductase family

Table 4: Probable recently duplicated genes

SAG0432	transcriptional regulator, AraC family; SAG0644 transcriptional regulator, AraC family
SAG0434	transposase, IS256 family, truncation; SAG0448 transposase, IS256 family
SAG0438	bacteriophage L54a, integrase, truncation; SAG1986 site-specific recombinase, phage integrase family; SAG1989 hypothetical protein; SAG1993 site-specific recombinase, phage integrase family; SAG2115 hypothetical protein
SAG0442	acetyltransferase, GNAT family; SAG0443 acetyltransferase, GNAT family
SAG0447	magnesium transporter, CorA family; SAG0875 magnesium transporter, CorA family, putative
SAG0508	beta-lactam resistance factor; SAG1349 beta-lactam resistance factor
SAG0566	prophage LambdaSal, single-strand binding protein; SAG1713 single-strand binding protein; SAG1863 prophage LambdaSa2, single-strand binding protein
SAG0603	conserved hypothetical protein; SAG1838 prophage LambdaSa2, holin, putative
SAG0604	prophage LambdaSal, lysin, putative; SAG1837 prophage LambdaSa2, lysin, putative
SAG0618	transposase OrfB, IS3 family, truncation; SAG0639 transposase OrfB, IS3 family; SAG1232 transposase OrfB, IS3 family, truncation; SAG1242 transposase OrfB, IS3 family, truncation
SAG0640	transposase OrfA, IS3 family; SAG1241 transposase OrfA, IS3 family
SAG0646	cell wall surface anchor family protein; SAG1404 cell wall surface anchor family protein

Table 4: Probable recently duplicated genes

SAG0647 sortase family protein; SAG0648 sortase family protein; SAG0650 sortase family protein
SAG0649 cell wall surface anchor family protein, putative; SAG1408 cell wall surface anchor family protein
SAG0676 proteinase, putative; SAG2053 serine protease, subtilase family, putative
SAG0679 protein of unknown function; SAG0680 protein of unknown function; SAG0681 conserved domain protein
SAG1002 protease, putative; SAG1465 protease, putative
SAG1025 hypothetical protein; SAG1033 FtsK/SpoIIIE family protein
SAG1067 IS861, transposase OrfA; SAG1526 IS861, transposase OrfA
SAG1068 IS861, transposase OrfB; SAG1256 IS861, transposase OrfB, truncation; SAG1527 IS861, transposase OrfB
SAG1140 conserved hypothetical protein; SAG1141 conserved hypothetical protein
SAG1164 glycosyl transferase CpsJ(V); SAG1165 glycosyl transferase CpsO(V)
SAG1182 phosphopentomutase; SAG2069 phosphopentomutase
SAG1225 conserved hypothetical protein; SAG1540 conserved hypothetical protein
SAG1228 ISSdy1, transposase OrfA; SAG1243 ISSdy1, transposase OrfA
SAG1229 ISSdy1, transposase OrfB; SAG1244 ISSdy1, transposase OrfB
SAG1253 transposase, ISL3 family; SAG2022 transposase, ISL3 family

Table 4: Probable recently duplicated genes

SAG1254 mercuric reductase; SAG2023 mercuric reductase
SAG1255 mercuric resistance operon regulatory protein MerR; SAG2024 mercuric resistance operon regulatory protein MerR
SAG1259 conserved hypothetical protein; SAG1272 conserved hypothetical protein
SAG1283 agglutinin receptor; SAG2021 cell wall surface anchor family protein
SAG1297 C-5 cytosine-specific DNA methylase; SAG1869 prophage LambdaSa2, type II DNA modification methyltransferase, putative
SAG1405 sortase family protein; SAG1406 sortase family protein
SAG1414 glycosyl transferase, group 2 family protein; SAG1415 glycosyl transferase, group 2 family protein
SAG1456 glycosyl transferase, family 8, degenerate; SAG2060 glycosyl transferase, family 8
SAG1521 transposase, IS30 family, putative; SAG1576 transposase, IS30 family, putative, truncation; SAG1795 transposase, IS30 family, putative
SAG1655 transcriptional regulator, MerR family; SAG1972 transcriptional regulator, MerR family
SAG1979 membrane protein, putative; SAG2034 membrane protein, putative
SAG1980 ABC transporter, ATP-binding protein; SAG2035 ABC transporter, ATP-binding protein
SAG1982 transcriptional regulator, Cro/C1 family; SAG2037 transcriptional regulator, Cro/C1 family
SAG1983 conserved hypothetical protein; SAG2039 conserved hypothetical protein

Table 4: Probable recently duplicated genes

SAG1984 conserved hypothetical protein TIGR00730; SAG2040 conserved hypothetical protein TIGR00730
SAG1988 conserved hypothetical protein; SAG2114 conserved hypothetical protein

Table 5

Strain	Source	Capsular serotype	Reference
090	Lancefield	Ia	
515	Houston	Ia	(1)
A909	Lancefield	Ia	(2)
Davis	Channing	Ia	
DK1	Houston	Ia	
DK8	Houston	Ia	
H36b	Lancefield	Ib	(2)
(S7) 7357b	Channing	Ib	(3)
18RS21	Lancefield	II	(4)
DK21	Houston	II	
COH1	Seattle	III	(5)
COH31	Seattle	III	(6)
D136C	Lancefield	III	(4)
M781	Houston	III	(7)
M732	Houston	III	(8)
1169NT1	Atlanta	V	(9)
2603V/R	Italy	V	This study
CJB111	Houston	V	(10)
JM9130013	Japan	VIII	(11)
SMU014	Japan	VIII	(11)
CJB110	Houston	Nontypeable	(12)

Table 5

1. Wessels, M. R., Paoletti, L. C., Rodewald, A. K., Michon, F., DiFabio, J., Jennings, H. J. & Kasper, D. L. (1993) *Infect Immun* **61**, 4760-6.
2. Wilkinson, H. W. & Eagon, R. G. (1971) *Infect Immun* **4**, 596-604.
3. Madoff, L. C., Michel, J. L., Gong, E. W., Rodewald, A. K. & Kasper, D. L. (1992) *Infect Immun* **60**, 4989-94.
4. Lancefield, R. C. (1975) in *New approaches for inducing natural immunity to pyogenic organisms* ed. Robbins, J. E. A. (National Institutes of Health, Bethesda, MD), pp. 145-151.
5. Wessels, M. R., Benedi, V.-J., Kasper, D. L., Heggen, L. M. & Rubens, C. E. (1991) in *Genetics and molecular biology of streptococci, lactococci, and enterococci* eds. Dunny, G. M., Cleary, P. P. & McKay, L. L. (American society for microbiology, Washington, DC), pp. 219-223.
6. Rubens, C. E., Wessels, M. R., Heggen, L. M. & Kasper, D. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7208-12.
7. Wessels, M. R., Paoletti, L. C., Kasper, D. L., DiFabio, J. L., Michon, F., Holme, K. & Jennings, H. J. (1990) *J Clin Invest* **86**, 1428-33.
8. Edwards, M. S., Wessels, M. R. & Baker, C. J. (1993) *Infect Immun* **61**, 2866-71.
9. Wilkinson, H. W. (1977) *J Clin Microbiol* **6**, 183-4.
10. Wessels, M. R., Paoletti, L. C., Pinel, J. & Kasper, D. L. (1995) *J Infect Dis* **171**, 879-84.
11. Lachenauer, C. S., Kasper, D. L., Shimada, J., Iciman, Y., Ohtsuka, H., Kaku, M., Paoletti, L. C. & Madoff, L. C. (1997) in *ICAAC*, pp. K-80.
12. Lachenauer, C. S., Creti, R., Michel, J. L. & Madoff, L. C. (2000) *Proc Natl Acad Sci USA* **97**, 9630-5.

Table 6

Cluster 1

SAG0230	conserved hypothetical protein
SAG0231	hypothetical protein
SAG0232	hypothetical protein
SAG0233	hypothetical protein
SAG0234	hypothetical protein
SAG0235	hypothetical protein

Cluster 2

SAG0222	conserved domain protein
SAG0223	conserved hypothetical protein, fusion
SAG0225	hypothetical protein
SAG0226	recombination protein
SAG0227	hypothetical protein
SAG0228	conserved hypothetical protein
SAG0229	conserved hypothetical protein

Cluster 3

SAG0634	hypothetical protein
SAG0635	acid phosphatase, class B
SAG0636	conserved hypothetical protein
SAG0638	cell wall surface anchor family protein, interruption-N
SAG0640	transposase OrfA, IS3 family

Table 6

SAG0642	hypothetical protein
SAG0643	chaperonin, 33 kDa, degenerate
SAG0644	transcriptional regulator, AraC family
SAG0645	cell wall surface anchor family protein
SAG0646	cell wall surface anchor family protein
SAG0647	sortase family protein
SAG0648	sortase family protein
SAG0649	cell wall surface anchor family protein, putative
SAG0650	sortase family protein
SAG0651	protein of unknown function

Cluster 4

SAG1898	PTS system, IID component
SAG1899	PTS system, IIC component
SAG1900	PTS system, IIB component
SAG1901	glucuronyl hydrolase
SAG1902	PTS system, IIA component
SAG1905	conserved hypothetical protein
SAG1906	carbohydrate kinase, PfkB family

Cluster 5

SAG0247	hypothetical protein
SAG0248	hypothetical protein

Table 6

SAG0249	hypothetical protein
SAG0674	hypothetical protein
SAG0675	putative secreted protein
SAG0676	proteinase, putative
SAG0677	hypothetical protein
SAG0680	protein of unknown function
SAG0681	conserved domain protein
SAG0684	ABC transporter, ATP-binding protein
SAG1698	conserved hypothetical protein

Cluster 6

SAG0261	IS1381, transposase OrfB
SAG0262	IS1381, transposase OrfA
SAG0965	IS1381, transposase OrfA
SAG0966	IS1381, transposase OrfB
SAG2002	IS1381, transposase OrfB

Cluster 7

SAG1027	conserved hypothetical protein
SAG1028	hypothetical protein
SAG1029	hypothetical protein
SAG1030	protein of unknown function
SAG1031	conserved domain protein

Table 6

SAG1032 conserved hypothetical protein

Cluster 8

SAG1253 transposase, ISL3 family

SAG1254 mercuric reductase

SAG1255 mercuric resistance operon regulatory protein MerR

SAG2022 transposase, ISL3 family

SAG2023 mercuric reductase

SAG2024 mercuric resistance operon regulatory protein MerR

Cluster 9

SAG1993 site-specific recombinase, phage integrase family

SAG1994 conserved hypothetical protein

SAG1995 hypothetical protein

SAG1996 cell wall surface anchor family protein, putative

SAG1997 hypothetical protein

SAG1998 hypothetical protein

SAG2000 membrane protein, putative

SAG2001 conjugal transfer protein, interruption-C

SAG2007 conserved hypothetical protein

SAG2008 conserved hypothetical protein

SAG2009 conserved hypothetical protein

SAG2010 hypothetical protein

Table 6

SAG2011	conserved hypothetical protein
SAG2012	hypothetical protein
SAG2016	hypothetical protein
SAG2017	transcriptional regulator, Cro/CI family
SAG2025	Mn ²⁺ /Fe ²⁺ transporter, NRAMP family

Cluster 10

SAG1039	conserved hypothetical protein
SAG1447	conserved hypothetical protein
SAG1448	glycosyl transferase, group 1 family protein
SAG1449	preprotein translocase SecA subunit, putative
SAG1450	conserved domain protein
SAG1452	conserved hypothetical protein
SAG1453	preprotein translocase SecY family protein
SAG1454	glycosyl transferase, putative
SAG1455	glycosyl transferase, group 2 family protein
SAG1456	glycosyl transferase, family 8, degenerate
SAG1459	glycosyl transferase family 8
SAG1460	glycosyl transferase, family 8
SAG1461	conserved hypothetical protein
SAG1462	cell wall surface anchor family protein
SAG1463	transcriptional regulator, RofA family, authentic point mutation
SAG1469	conserved hypothetical protein

Table 6

SAG1471	conserved hypothetical protein
SAG1933	PTS system, IIC component, putative

Cluster 11

SAG0009	hypothetical protein
SAG0120	hypothetical protein
SAG0157	deoxyribonuclease-related protein, degenerate
SAG0186	hypothetical protein
SAG0216	hypothetical protein
SAG0236	hypothetical protein
SAG0307	hypothetical protein
SAG0308	ABC transporter, ATP-binding protein
SAG0311	DNA-binding response regulator, authentic point mutation
SAG0518	peptide chain release factor 2, programmed frameshift
SAG0553	hypothetical protein
SAG0555	prophage LambdaSa1, antirepressor, putative
SAG0564	conserved hypothetical protein
SAG0579	conserved hypothetical protein
SAG0580	conserved hypothetical protein, truncation
SAG0611	transposase, degenerate
SAG0637	transcriptional regulator, TetR family, putative, authentic frameshift
SAG0641	Tn5252, Orf 10 protein, degenerate
SAG0652	Tn5252, Orf 28 protein, degenerate

Table 6

SAG0655	conserved hypothetical protein
SAG0678	endopeptidase O, degenerate
SAG0683	transmembrane protein Vexp3, putative, degenerate
SAG0855	glycogen biosynthesis protein GlgD, authentic frameshift
SAG0898	hypothetical protein
SAG0899	hypothetical protein
SAG0901	hypothetical protein
SAG0902	hypothetical protein
SAG0903	hypothetical protein
SAG0917	Tn916, hypothetical protein
SAG0920	Tn916, hypothetical protein
SAG0922	Tn916, hypothetical protein
SAG0924	Tn916, tetM leader peptide
SAG0928	Tn916, hypothetical protein, authentic frameshift
SAG0936	Tn916, hypothetical protein
SAG0943	hypothetical protein
SAG0972	conserved hypothetical protein, authentic frameshift
SAG1023	hypothetical protein
SAG1080	hypothetical protein
SAG1123	hypothetical protein
SAG1129	hypothetical protein
SAG1136	conserved hypothetical protein
SAG1217	conserved hypothetical protein, authentic frameshift

Table 6

SAG1231	transposase OrfB, IS3 family, degenerate
SAG1242	transposase OrfB, IS3 family, truncation
SAG1309	hypothetical protein
SAG1331	R5 protein
SAG1437	hypothetical protein
SAG1445	MutT/nudix family protein, authentic frameshift
SAG1484	ribosomal protein L33
SAG1493	hypothetical protein
SAG1539	hypothetical protein
SAG1543	conserved hypothetical protein, authentic frameshift
SAG1560	hypothetical protein
SAG1568	phosphoserine aminotransferase, authentic frameshift
SAG1570	conserved hypothetical protein
SAG1601	conserved hypothetical protein
SAG1644	hypothetical protein
SAG1646	hypothetical protein
SAG1699	hypothetical protein
SAG1705	peptidase, M24 family, authentic point mutation
SAG1708	hypothetical protein
SAG1857	prophage LambdaSa2, HNH endonuclease family protein
SAG1864	hypothetical protein
SAG1868	hypothetical protein

Table 6

SAG1869	prophage LambdaSa2, type II DNA modification methyltransferase, putative
SAG1872	hypothetical protein
SAG1874	hypothetical protein
SAG1876	prophage LambdaSa2, HNH endonuclease family protein
SAG1878	conserved domain protein
SAG1881	hypothetical protein
SAG1883	conserved hypothetical protein
SAG1886	hypothetical protein
SAG1903	hypothetical protein
SAG1937	streptococcal histidine triad family protein, degenerate
SAG1971	hypothetical protein
SAG1979	membrane protein, putative
SAG1980	ABC transporter, ATP-binding protein
SAG1981	hypothetical protein
SAG1982	transcriptional regulator, Cro/CI family
SAG1983	conserved hypothetical protein
SAG1984	conserved hypothetical protein TIGR00730
SAG1985	hypothetical protein
SAG1991	transcriptional regulator, Cro/CI family
SAG1992	protein of unknown function
SAG1999	hypothetical protein
SAG2004	conjugal transfer protein, interruption-N

Table 6

SAG2039	conserved hypothetical protein
SAG2044	hypothetical protein
SAG2052	hypothetical protein
SAG2065	ribosomal protein L33
SAG2094	competence/damage-inducible protein CinA, authentic frameshift
SAG2099	hypothetical protein

Cluster 12

SAG1164	glycosyl transferase CpsJ(V)
SAG1165	glycosyl transferase CpsO(V)
SAG1166	glycosyl transferase CpsN(V)
SAG1167	polysaccharide biosynthesis protein CpsM(V)
SAG1168	polysaccharide biosynthesis protein cpsH(V)

Cluster 13

SAG0581	conserved hypothetical protein
SAG0582	conserved hypothetical protein
SAG0583	conserved hypothetical protein
SAG0585	conserved hypothetical protein
SAG0586	conserved hypothetical protein
SAG0587	prophage LambdaSa1, structural protein, putative
SAG0588	conserved hypothetical protein
SAG0589	conserved hypothetical protein

Table 6

SAG0590	conserved hypothetical protein
SAG0591	conserved hypothetical protein
SAG0593	prophage LambdaSa1, structural protein
SAG0594	conserved hypothetical protein
SAG0595	conserved hypothetical protein
SAG0596	prophage LambdaSa1, pblA protein, internal deletion

Cluster 14

SAG0915	Tn916, transposase
SAG0918	Tn916, hypothetical protein
SAG0919	Tn916, hypothetical protein
SAG0921	Tn916, transcriptional regulator, putative
SAG0925	Tn916, hypothetical protein
SAG0926	Tn916, NLP/P60 family protein
SAG0927	membrane protein, putative
SAG0929	Tn916, hypothetical protein
SAG0930	Tn916, hypothetical protein
SAG0931	Tn916, hypothetical protein
SAG0932	Tn916, transcriptional regulator, putative
SAG0933	Tn916, FtsK/SpoIIIE family protein
SAG0934	Tn916, hypothetical protein
SAG0935	Tn916, hypothetical protein
SAG0937	ABC transporter, ATP-binding protein, authentic frameshift

Table 6

Cluster 15

SAG1835	conserved hypothetical protein
SAG1837	prophage LambdaSa2, lysin, putative
SAG1839	conserved hypothetical protein
SAG1840	hypothetical protein
SAG1842	prophage LambdaSa2, PblB, putative
SAG1843	conserved hypothetical protein
SAG1844	conserved hypothetical protein
SAG1849	hypothetical protein
SAG1851	conserved domain protein
SAG1852	conserved domain protein
SAG1853	prophage LambdaSa2, protease, putative
SAG1854	conserved hypothetical protein
SAG1855	prophage LambdaSa2, terminase large subunit, putative
SAG1856	hypothetical protein
SAG1858	hypothetical protein
SAG1859	prophage LambdaSa2, site-specific recombinase, phage integrase family
SAG1860	conserved hypothetical protein
SAG1861	prophage LambdaSa2, transcriptional regulator, Cro/CI family
SAG1862	hypothetical protein
SAG1863	prophage LambdaSa2, single-strand binding protein
SAG1865	conserved hypothetical protein

Table 6

SAG1866	conserved hypothetical protein
SAG1867	conserved hypothetical protein
SAG1870	prophage LambdaSa2, DNA replication protein DnaC, putative
SAG1871	prophage LambdaSa2, bacteriophage replication protein/hypothetical protein, truncation/fusion
SAG1873	prophage LambdaSa2, replicative DNA helicase
SAG1877	prophage LambdaSa2, antirepressor protein, putative
SAG1879	hypothetical protein
SAG1882	prophage LambdaSa2, repressor protein, putative
SAG1884	hypothetical protein
SAG1885	prophage LambdaSa2, site-specific recombinase, phage integrase family

Cluster 16

SAG1247	site-specific recombinase, phage integrase family
SAG1250	Tn5252, relaxase
SAG1251	Tn5252, Orf 9 protein
SAG1252	Tn5252, Orf 10 protein
SAG1256	IS861, transposase OrfB, truncation
SAG1257	cation-transporting ATPase, E1-E2 family
SAG1258	cadmium efflux system accessory protein
SAG1259	conserved hypothetical protein
SAG1260	hypothetical protein
SAG1261	conserved hypothetical protein

Table 6

SAG1262	cation-transporting ATPase, E1-E2 family
SAG1263	conserved domain protein, authentic frameshift
SAG1264	transcriptional repressor CopY, putative
SAG1265	cadmium resistance transporter, putative
SAG1266	hypothetical protein
SAG1267	hypothetical protein
SAG1268	repressor protein, putative
SAG1270	ImpB/MucB/SamB family protein
SAG1271	conserved hypothetical protein
SAG1272	conserved hypothetical protein
SAG1273	conserved hypothetical protein
SAG1274	conserved hypothetical protein
SAG1276	conserved hypothetical protein
SAG1277	hypothetical protein
SAG1278	hypothetical protein
SAG1279	conserved domain protein
SAG1280	SNF2 family protein
SAG1281	hypothetical protein
SAG1283	agglutinin receptor
SAG1284	abortive infection protein AbiGI
SAG1285	abortive infection protein AbiGII
SAG1286	Tn5252, Orf28
SAG1287	Tn5252, Orf26

Table 6

SAG1288	Tn5252, Orf25, degenerate
SAG1289	Tn5252, Orf23
SAG1290	hypothetical protein
SAG1291	Tn5252, Orf 21 protein, internal deletion
SAG1292	hypothetical protein
SAG1293	protease, putative
SAG1294	conserved hypothetical protein
SAG1295	conserved hypothetical protein
SAG1296	conserved hypothetical protein
SAG1297	C-5 cytosine-specific DNA methylase
SAG1299	conserved hypothetical protein
SAG1304	hypothetical protein

Table 7

Locus	Annotation
Housekeeping	
SAG0466	thiolase
SAG0471	glucokinase
SAG0492	amino acid ABC transporter, ATP-binding protein
SAG0767	D-alanine--D-alanine ligase
SAG1086	xanthine phosphoribosyltransferase
SAG1600	glutamate racemase
SAG1680	shikimate 5-dehydrogenase
SAG1723	signal peptidase I
Surface-exposed	
SAG0079	adenylate kinase
SAG0093	D-alanyl-D-alanine carboxypeptidase family protein
SAG0163	competence protein CglA
SAG0290	ABC transporter, substrate-binding protein
SAG0368	protein of unknown function
SAG0503	lipase/acylhydrolase
SAG1473	cell wall surface anchor family protein
SAG1552	conserved hypothetical protein
SAG1641	YaeC family protein
SAG2147	protein of unknown function/lipoprotein, putative
SAG2148	LysM domain protein

Table 8: GBS genes shared with GAS and pn umococcus

ORFxxxxx Ann tation

ORF00003	PcsB protein (pscB)
ORF00004	ribose-phosphate pyrophosphokinase (prsA)
ORF00005	aminotransferase, class I
ORF00006	recombination protein O
ORF00009	fatty acid/phospholipid synthesis protein PlsX (plsX)
ORF00011	phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)
ORF00012	phosphoribosylformylglycinamide synthase, putative
ORF00013	amidophosphoribosyltransferase (purF)
ORF00014	phosphoribosylformylglycinamide cyclo-ligase (purM)
ORF00015	phosphoribosylglycinamide formyltransferase (purN)
ORF00020	group B streptococcal surface immunogenic protein
ORF00021	N-acetylmannosamine-6-P epimerase, putative
ORF00022	sugar ABC transporter, sugar-binding protein
ORF00023	sugar ABC transporter, permease protein
ORF00024	sugar ABC transporter, permease protein
ORF00026	conserved hypothetical protein
ORF00027	N-acetylneuraminate lyase, putative
ORF00028	expressed ROK family protein
ORF00030	phosphosugar-binding transcriptional regulator, RpiR family, putative
ORF00031	phosphoribosylamine--glycine ligase (purD)
ORF00032	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)
ORF00033	phosphoribosylaminoimidazole carboxylase, ATPase subunit (purK)
ORF00036	adenylosuccinate lyase (purB)
ORF00037	transcriptional regulator, Cro/Ci family
ORF00038	Holliday junction DNA helicase RuvB (ruvB)
ORF00039	phosphotyrosine protein phosphatase, low molecular weight
ORF00040	MORN motif family protein
ORF00041	membrane protein, putative
ORF00043	alcohol dehydrogenase, propanol-preferring (adhP)
ORF00045	MATE efflux family protein
ORF00046	ribosomal protein S10 (rpsJ)
ORF00047	ribosomal protein L3 (rplC)
ORF00048	ribosomal protein L4 (rplD)
ORF00049	ribosomal protein L23 (rplW)
ORF00050	ribosomal protein L2 (rplB)
ORF00052	ribosomal protein S19 (rpsS)
ORF00054	ribosomal protein L22 (rplV)
ORF00055	ribosomal protein S3 (rpsC)
ORF00056	ribosomal protein L16 (rplP)
ORF00058	ribosomal protein L29 (rpmC)
ORF00059	ribosomal protein S17 (rpsQ)
ORF00060	ribosomal protein L14 (rplN)
ORF00061	ribosomal protein L24 (rplX)
ORF00063	ribosomal protein L5 (rplE)
ORF00065	ribosomal protein S8 (rpsH)
ORF00066	ribosomal protein L6 (rplF)
ORF00068	ribosomal protein L18 (rplR)
ORF00069	ribosomal protein S5 (rpsE)
ORF00070	ribosomal protein L30 (rpmD)
ORF00071	ribosomal protein L15 (rplO)
ORF00072	preprotein translocase, SecY subunit
ORF00073	adenylate kinase (adk)
ORF00074	translation initiation factor IF-1 (InfA)
ORF00075	ribosomal protein L36 (rpmJ)
ORF00077	ribosomal protein S13 (rpsM)

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00078	ribosomal protein S11 (rpsK)
ORF00080	DNA-directed RNA polymerase, alpha subunit (rpoA)
ORF00093	transcriptional regulator ComX1, putative
ORF00094	phosphoglycerate mutase family protein
ORF00097	heat-inducible transcription repressor HrcA (hrcA)
ORF00098	heat shock protein GrpE (grpE)
ORF00099	dnaK protein (dnaK)
ORF00100	dnaJ protein (dnaJ)
ORF00101	transcriptional regulator, GntR family
ORF00102	tRNA pseudouridine synthase A (truA)
ORF00103	phosphomethylpyrimidine kinase, putative
ORF00104	conserved hypothetical protein
ORF00105	conserved hypothetical protein
ORF00106	conserved hypothetical protein
ORF00107	trigger factor (tig)
ORF00108	DNA-directed RNA polymerase, delta subunit, putative
ORF00109	CTP synthase (pyrG)
ORF00111	deoxyuridine 5'-triphosphate nucleotidohydrolase (dut)
ORF00113	carbonic anhydrase-related protein
ORF00115	pyridine nucleotide-disulphide oxidoreductase family protein
ORF00116	glutamyl-tRNA synthetase (glbX)
ORF00119	ribose ABC transporter, ATP-binding protein (rbsA)
ORF00122	ribose operon repressor RbsR (rbsR)
ORF00125	ABC transporter, ATP-binding protein
ORF00126	DNA-binding response regulator
ORF00128	sensor histidine kinase
ORF00131	fructose-bisphosphate aldolase (fba)
ORF00132	L-2-hydroxyisocaproate dehydrogenase
ORF00133	ribosomal protein L28 (rpmB)
ORF00134	conserved hypothetical protein
ORF00135	DAK2 domain protein
ORF00136	expressed SPFH domain/Band 7 family protein
ORF00141	amino acid ABC transporter, ATP-binding protein
ORF00142	amino acid ABC transporter, amino acid-binding protein/permease protein
ORF00143	conserved hypothetical protein
ORF00145	undecaprenol kinase, putative
ORF00146	negative regulator of competence MecA, putative
ORF00149	ABC transporter, ATP-binding protein
ORF00150	conserved hypothetical protein
ORF00151	selenocysteine lyase (csdB)
ORF00152	NifU family protein
ORF00153	conserved hypothetical protein
ORF00155	D-alanyl-D-alanine carboxypeptidase
ORF00158	oligopeptide ABC transporter, permease protein
ORF00160	oligopeptide ABC transporter, ATP-binding protein
ORF00161	oligopeptide ABC transporter, ATP-binding protein
ORF00167	adc operon repressor AdcR (adcR)
ORF00168	zinc ABC transporter, ATP-binding protein
ORF00169	zinc ABC transporter, permease protein
ORF00172	tyrosyl-tRNA synthetase (tyrS)
ORF00173	penicillin-binding protein 1B, putative
ORF00174	DNA-directed RNA polymerase, beta subunit (rpoB)
ORF00176	DNA-directed RNA polymerase beta' subunit (rpoC)
ORF00178	conserved hypothetical protein
ORF00179	competence protein CglA (cglA)

Table 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00180	competence protein CglB (cglB)
ORF00181	conserved hypothetical protein
ORF00183	conserved hypothetical protein
ORF00184	acetate kinase (ackA)
ORF00190	pyrroline-5-carboxylate reductase (proC)
ORF00191	glutamyl-aminopeptidase (pepA)
ORF00198	single-strand binding protein (ssb)
ORF00211	PTS system, IIBC components
ORF00212	alpha amylase family protein
ORF00214	transcriptional antiterminator, BglG family
ORF00219	PTS system, IIC component, putative
ORF00224	ribosomal protein S15 (rpsO)
ORF00225	polyribonucleotide nucleotidyltransferase (pnp)
ORF00227	serine O-acetyltransferase (cysE)
ORF00229	cysteinyI-tRNA synthetase (cysS)
ORF00230	conserved hypothetical protein
ORF00231	RNA methyltransferase, TrmH family, group 3
ORF00233	DegV family protein
ORF00236	ribosomal protein L13 (rplM)
ORF00237	ribosomal protein S9 (rpsI)
ORF00261	transcriptional regulator MutR family
ORF00262	transporter, putative
ORF00263	amino acid ABC transporter, permease protein
ORF00264	amino acid ABC transporter, amino acid-binding protein
ORF00265	amino acid ABC transporter, permease protein
ORF00266	amino acid ABC transporter, ATP-binding protein
ORF00295	N-acetylglucosamine-6-phosphate deacetylase (nagA)
ORF00296	conserved hypothetical protein
ORF00297	glycyl-tRNA synthetase, alpha subunit (glyQ)
ORF00299	glycyl-tRNA synthetase, beta subunit (glyS)
ORF00300	conserved hypothetical protein
ORF00302	glycerol kinase (glpK)
ORF00303	alpha-glycerophosphate oxidase
ORF00304	glycerol uptake facilitator protein (glpF)
ORF00306	conserved hypothetical protein
ORF00307	transketolase (tkt)
ORF00309	ABC transporter, ATP-binding protein
ORF00310	membrane protein, putative
ORF00313	PTS system, IIBC components
ORF00314	glutamate 5-kinase (proB)
ORF00315	gamma-glutamyl phosphate reductase (proA)
ORF00316	conserved hypothetical protein TIGR00006
ORF00318	penicillin-binding protein 2X (pbpX)
ORF00319	phospho-N-acetylmuramoyl-pentapeptide-transferase (mraY)
ORF00320	ATP-dependent RNA helicase, DEAD/DEAH box family
ORF00321	ABC transporter, substrate-binding protein
ORF00322	amino acid ABC transporter, permease protein
ORF00323	amino acid ABC transporter, ATP-binding protein
ORF00325	thioredoxin reductase (trxB)
ORF00326	conserved hypothetical protein
ORF00327	NAD synthetase (nadE)
ORF00328	aminopeptidase C (pepC)
ORF00329	penicillin-binding protein 1A (pbp1A)
ORF00330	recombination protein U (recU)
ORF00331	conserved hypothetical protein

Table 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00335 conserved hypothetical protein
ORF00336 conserved hypothetical protein
ORF00337 autoinducer-2 production protein LuxS (luxS)
ORF00338 KH domain protein
ORF00348 guanylate kinase (gmk)
ORF00349 DNA-directed RNA polymerase, omega subunit, putative
ORF00350 primosomal protein N' (priA)
ORF00351 methionyl-tRNA formyltransferase (fmt)
ORF00352 Sun protein (sun)
ORF00353 serine/threonine phosphatase, putative
ORF00354 serine/threonine protein kinase
ORF00355 conserved hypothetical protein
ORF00356 sensor histidine kinase, putative
ORF00358 DNA-binding response regulator
ORF00359 hydrolase, haloacid dehalogenase family/peptidyl-prolyl cis-trans isomerase, cyclophilin type
ORF00360 general stress protein, putative
ORF00361 pyruvate formate-lyase-activating enzyme (pflA)
ORF00362 transcriptional regulator, DeoR family
ORF00363 transcriptional regulator, putative
ORF00364 PTS system, cellobiose-specific IIA component (celC)
ORF00366 PTS system, cellobiose-specific IIB component (celA)
ORF00367 PTS system, cellobiose-specific IIC component (celB)
ORF00368 formate acetyltransferase (pflD)
ORF00369 transaldolase family protein
ORF00371 glycerol dehydrogenase (gldA)
ORF00372 cysteine synthase A (cysK)
ORF00373 conserved hypothetical protein TIGR00257
ORF00374 helicase, putative
ORF00375 competence protein F, putative
ORF00376 ribosomal subunit interface protein (yfiA)
ORF00385 enoyl-CoA hydratase/isomerase family protein
ORF00386 transcriptional regulator, MarR family
ORF00387 3-oxoacyl-(acyl-carrier-protein) synthase III (fabH)
ORF00388 acyl carrier protein (acpP)
ORF00390 enoyl-(acyl-carrier-protein) reductase II (fabK)
ORF00391 malonyl CoA-acyl carrier protein transacylase (fabD)
ORF00392 3-oxoacyl-(acyl-carrier-protein) reductase (fabG)
ORF00393 3-oxoacyl-(acyl-carrier-protein) synthase II (fabF)
ORF00394 acetyl-CoA carboxylase, biotin carboxyl carrier protein (accB)
ORF00395 (3R)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase (fabZ)
ORF00396 acetyl-CoA carboxylase, biotin carboxylase (accC)
ORF00397 acetyl-CoA carboxylase, carboxyl transferase, beta subunit (accD)
ORF00398 acetyl-CoA carboxylase, carboxyl transferase, alpha subunit (accA)
ORF00400 seryl-tRNA synthetase (serS)
ORF00403 conserved hypothetical protein
ORF00404 PTS system, mannose-specific IID component
ORF00405 PTS system, mannose-specific IIC component (manM)
ORF00406 PTS system, mannose-specific IIB components (manL)
ORF00407 hydrolase, haloacid dehalogenase-like family
ORF00410 xanthine/uracil permease family protein
ORF00411 conserved hypothetical protein TIGR00150, putative
ORF00412 acetyltransferase, GNAT family
ORF00413 expressed protein of unknown function
ORF00415 HIT family protein (hit)
ORF00419 ABC transporter, ATP-binding protein

Table 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00421	ABC transporter, permease protein
ORF00422	conserved hypothetical protein
ORF00423	conserved hypothetical protein TIGR00091
ORF00424	conserved hypothetical protein, POINT MUTATION
ORF00425	N utilization substance protein A (nusa)
ORF00426	conserved hypothetical protein
ORF00427	ribosomal protein L7A family
ORF00428	translation initiation factor IF-2
ORF00429	ribosome-binding factor A (rbfA)
ORF00432	copper-transporter ATPase CopA
ORF00435	hydrolase, haloacid dehalogenase-like family
ORF00436	DNA polymerase I (polA)
ORF00437	CoA binding domain protein
ORF00440	DNA-binding response regulator
ORF00441	sensor histidine kinase
ORF00443	queuine tRNA-ribosyltransferase (tgt)
ORF00444	conserved hypothetical protein
ORF00449	glucose-6-phosphate isomerase (pgi)
ORF00451	rhomboid family protein
ORF00452	expressed putative lipoprotein
ORF00453	UTP-glucose-1-phosphate uridylyltransferase (galU)
ORF00454	glycerol-3-phosphate dehydrogenase (NAD(P)+) (gpsA)
ORF00455	ribonuclease P protein component (mpA)
ORF00456	SpoIIJ family protein
ORF00458	R3H domain protein
ORF00463	conserved hypothetical protein
ORF00464	RecX protein
ORF00465	RNA methyltransferase, TrmA family
ORF00470	ribonucleoside-diphosphate reductase 2, beta subunit (nrdF)
ORF00472	ribonucleoside-diphosphate reductase 2, alpha subunit (nrdE)
ORF00482	alcohol dehydrogenase, zinc-containing
ORF00483	oxidoreductase, aldo/keto reductase family
ORF00484	cation efflux system protein
ORF00485	transcriptional regulator, TetR family
ORF00496	conserved hypothetical protein
ORF00500	acetyltransferase, GNAT family
ORF00501	conserved hypothetical protein
ORF00502	valyl-tRNA synthetase (valS)
ORF00508	aspartate-ammonia ligase (asnA)
ORF00511	type II DNA modification methyltransferase, putative
ORF00513	phosphopantetheine adenylyltransferase (coaD)
ORF00515	conserved hypothetical protein
ORF00519	conserved hypothetical protein
ORF00520	conserved hypothetical protein TIGR00048
ORF00522	ABC transporter, ATP-binding/permease protein
ORF00523	ABC transporter, ATP-binding/permease protein
ORF00524	anthranilate synthase component II (trpG)
ORF00532	endonuclease III (nth)
ORF00534	conserved hypothetical protein
ORF00535	glucokinase (glk)
ORF00536	expressed protein with rhodanese domain
ORF00537	elongation factor Tu family protein
ORF00540	UDP-N-acetylmuramoylalanine-D-glutamate ligase (murD)
ORF00541	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (murG)

Tabl 8: GBS g n s shared with GAS and pneumococ us

ORFxxxxx Annotati n

ORF00542 cell division protein DivIB, putative
ORF00544 cell division protein FtsA (ftsA)
ORF00545 cell division protein FtsZ (ftsZ)
ORF00546 ylmE protein, putative
ORF00547 ylmF protein (ylmF)
ORF00549 ylmH protein (ylmH)
ORF00550 cell division protein DivIVA, putative
ORF00552 isoleucyl-tRNA synthetase (ileS)
ORF00553 conserved hypothetical protein
ORF00554 MutT/nudix family protein
ORF00555 ATP-dependent Clp protease, ATP-binding subunit
ORF00557 conserved hypothetical protein
ORF00558 amino acid ABC transporter, permease protein
ORF00559 amino acid ABC transporter, ATP-binding protein
ORF00560 phosphoglucomutase/phosphomannomutase family protein
ORF00562 methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase (folD)
ORF00564 exodeoxyribonuclease VII, large subunit (xseA)
ORF00566 geranyltransferase, putative
ORF00567 hemolysin A
ORF00570 DNA repair protein RecN (recN)
ORF00571 expressed DegV family protein
ORF00574 DNA-binding protein HU (hup)
ORF00576 dihydroorotate dehydrogenase A (pyrDA)
ORF00577 beta-lactam resistance factor (fibB)
ORF00578 beta-lactam resistance factor (fibA)
ORF00579 murM protein, putative
ORF00580 hydrolase, haloacid dehalogenase-like family
ORF00581 HD domain protein
ORF00582 conserved hypothetical protein
ORF00583 cation-transporting ATPase, E1-E2 family
ORF00588 cell division ABC transporter, ATP-binding protein FtsE (ftsE)
ORF00589 cell division ABC transporter, permease protein FtsX (ftsX)
ORF00591 metallo-beta-lactamase superfamily protein
ORF00593 DNA polymerase III, epsilon subunit/ATP-dependent helicase DinG
ORF00595 aspartate aminotransferase (aspC)
ORF00596 asparaginyl-tRNA synthetase (asnS)
ORF00601 conserved hypothetical protein
ORF00602 conserved hypothetical protein
ORF00603 conserved hypothetical protein
ORF00605 zinc ABC transporter, zinc-binding adhesion liprotein
ORF00606 ribosomal protein L31 (rpmE)
ORF00607 DHH family protein
ORF00609 flavodoxin
ORF00614 ribosomal protein L19 (rplS)
ORF00640 prophage LambdaSa1, single-strand binding protein (ssb)
ORF00693 DNA-binding response regulator VncR (vncR)
ORF00694 sensor histidine kinase VncS (vncS)
ORF00699 rod shape-determining protein RodA, putative (rodA)
ORF00700 hydrolase, haloacid dehalogenase-like family
ORF00701 DNA gyrase, B subunit (gyrB)
ORF00702 septation ring formation regulator EzrA, putative
ORF00705 conserved hypothetical protein
ORF00706 enolase (eno)
ORF00708 3-phosphoshikimate 1-carboxyvinyltransferase (aroA)
ORF00709 shikimate kinase (aroK)

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00710	psr protein
ORF00711	RNA methyltransferase, TrmA family
ORF00729	sortase family protein
ORF00731	sortase family protein
ORF00734	sortase family protein, FRAMESHIFT
ORF00743	ABC transporter, ATP-binding protein
ORF00744	membrane protein
ORF00745	conserved hypothetical protein
ORF00748	cylG protein (cylG)
ORF00776	DNA-entry nuclease, putative
ORF00789	2-keto-3-deoxygluconate kinase
ORF00792	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase (eda)
ORF00798	proline dipeptidase (pepQ)
ORF00799	transcriptional regulator, RegM family
ORF00802	glycosyl transferase, group 1 family protein
ORF00803	threonyl-tRNA synthetase (thrS)
ORF00804	DNA-binding response regulator
ORF00808	amino acid ABC transporter, permease protein
ORF00810	amino acid ABC transporter, ATP-binding protein
ORF00811	DNA-binding response regulator
ORF00812	sensory box histidine kinase
ORF00813	metallo-beta-lactamase family protein
ORF00815	ribonuclease III (rnc)
ORF00816	expressed putative chromosome segregation SMC protein
ORF00817	hydrolase, haloacid dehalogenase-like family
ORF00818	hydrolase, haloacid dehalogenase-like family
ORF00819	signal recognition particle-docking protein FtsY (ftsY)
ORF00820	ABC transporter, substrate-binding protein
ORF00821	ABC transporter, permease protein, putative
ORF00824	transcriptional accessory protein Tex, putative
ORF00825	conserved hypothetical protein
ORF00828	HPr(Ser) kinase/phosphatase (hprK)
ORF00830	prolipoprotein diacylglycerol transferase (lgt)
ORF00832	conserved hypothetical protein
ORF00835	peptidase, U32 family, putative
ORF00836	peptidase, U32 family
ORF00837	conserved hypothetical protein
ORF00844	lysyl-tRNA synthetase (lysS)
ORF00846	phosphoglycerate mutase family protein
ORF00847	ebsC family protein, putative
ORF00850	peptidase, U32 family
ORF00855	oligoendopeptidase F, putative
ORF00856	phosphoenolpyruvate carboxylase (ppc)
ORF00859	cell division protein, FtsW/RodA/SpoVE family (ftsW)
ORF00861	translation elongation factor Tu (tuf)
ORF00863	triosephosphate isomerase (tpiA)
ORF00865	phosphoglycerate mutase (gpmA)
ORF00867	recombination protein RecR (recR)
ORF00868	D-alanine-D-alanine ligase
ORF00869	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase (murF)
ORF00870	oxalate:formate antiporter
ORF00871	membrane protein, putative
ORF00873	peptide chain release factor 3 (prfC)
ORF00876	ABC transporter, ATP-binding protein
ORF00880	ATP-dependent RNA helicase, DEAD/DEAH box family

Table 8: GBS genes shared with GAS and *pn umococcus*

ORFxxxxx Annotation

ORF00882 conserved hypothetical protein
ORF00883 conserved hypothetical protein
ORF00884 acyltransferase family protein
ORF00885 competence protein CeiA (ceiA)
ORF00887 DNA internalization-related competence protein ComEC/Rec2
ORF00889 sugar-binding transcriptional regulator, LacI family
ORF00892 DNA polymerase III, delta subunit, putative
ORF00893 superoxide dismutase, Fe-Mn (sodA)
ORF00894 transcriptional antiterminator LicT
ORF00895 PTS system, beta-glucosides-specific IIABC components
ORF00896 6-phospho-beta-glucosidase (bglA)
ORF00899 glycerate kinase 2 (garK)
ORF00904 S-adenosylmethionine:tRNA ribosyltransferase-isomerase (queA)
ORF00906 glucosamine-6-phosphate isomerase (nagB)
ORF00908 ribosomal small subunit pseudouridine synthase
ORF00911 competence protein CoiA (coiA)
ORF00912 oligoendopeptidase B (pepB)
ORF00914 O-methyltransferase family protein
ORF00916 protease maturation protein, putative
ORF00919 alanyl-tRNA synthetase (alaS)
ORF00925 transcriptional regulator, Cro/C1 family
ORF00928 ribonucleoside-diphosphate reductase 2, beta subunit (nrdF)
ORF00929 ribonucleoside-diphosphate reductase 2, alpha subunit (nrdE)
ORF00930 ribonucleoside-diphosphate reductase 2, NrdH-redoxin (nrdH)
ORF00931 phosphocarrier protein HPr (ptsH)
ORF00932 phosphoenolpyruvate-protein phosphotransferase (ptsI)
ORF00933 glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent (gapN)
ORF00934 polysaccharide deacetylase family protein
ORF00935 ATP-dependent RNA helicase, DEAD/DEAH box family
ORF00936 uridine kinase (udk)
ORF00937 conserved hypothetical protein
ORF00938 DNA polymerase III, gamma and tau subunits (dnaX)
ORF00940 biotin--acetyl-CoA-carboxylase ligase
ORF00941 S-adenosylmethionine synthetase (metK)
ORF00955 UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA)
ORF00956 acetyltransferase, GNAT family
ORF00957 CBS domain protein
ORF00958 methionine aminopeptidase, type I (map)
ORF00959 ribonuclease BN, putative
ORF00962 conserved hypothetical protein
ORF00963 DNA ligase, NAD-dependent (ligA)
ORF00964 BmrU protein, putative
ORF00966 pullulanase, putative
ORF00973 ATP synthase F0, A subunit (atpB)
ORF00974 ATP synthase F0, B subunit (atpF)
ORF00975 ATP synthase F1, delta subunit (atpH)
ORF00976 ATP synthase F1, alpha subunit (atpA)
ORF00977 ATP synthase F1, gamma subunit (atpG)
ORF00978 ATP synthase F1, beta subunit (atpD)
ORF00979 ATP synthase F1, epsilon subunit (atpC)
ORF00981 UDP-N-acetylglucosamine 1-carboxyvinyltransferase (murA)
ORF00983 DNA-entry nuclease (endA)
ORF00984 phenylalanyl-tRNA synthetase, alpha subunit (pheS)
ORF00986 phenylalanyl-tRNA synthetase, beta subunit (pheT)
ORF00988 exonuclease RxB (rxB)

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF00989	exonuclease RexA (rexA)
ORF00991	tRNA modification GTPase TrmE (trmE)
ORF00992	ABC transporter, ATP-binding protein
ORF00993	acetoin dehydrogenase, thymine PPI dependent, E1 component, alpha subunit
ORF00994	acetoin dehydrogenase, thymine PPI dependent, E1 component, beta subunit
ORF00995	acetoin dehydrogenase, thymine PPI dependent, E2 component, dihydrolipoamide
ORF00996	acetoin dehydrogenase, thymine PPI dependent, E3 component, dihydrolipoamide dehydrogenase
ORF00997	lipoate-protein ligase A (lplA)
ORF00998	cobyrinic acid synthase, putative
ORF00999	mur ligase family protein
ORF01000	conserved hypothetical protein TIGR00159
ORF01001	expressed protein of unknown function
ORF01002	phosphoglucomutase/phosphomannomutase family protein
ORF01005	oxygen-independent coproporphyrinogen III oxidase, putative
ORF01006	conserved hypothetical protein
ORF01007	hydrolase, haloacid dehalogenase-like family
ORF01008	conserved hypothetical protein
ORF01023	GTP-binding protein LepA (lepA)
ORF01027	PilB-related protein
ORF01030	cation-transporting ATPase, E1-E2 family
ORF01033	conserved hypothetical protein
ORF01040	Tn916, tetracycline resistance protein (tetM)
ORF01057	transcriptional regulator, GntR family
ORF01058	DNA polymerase III, alpha subunit (dnaE)
ORF01059	6-phosphofructokinase (pfk)
ORF01060	pyruvate kinase (pyk)
ORF01063	glucosamine-fructose-6-phosphate aminotransferase (isomerizing) (glmS)
ORF01066	phnA protein (phnA)
ORF01068	amino acid ABC transporter, permease protein
ORF01069	amino acid ABC transporter, ATP-binding protein
ORF01070	amino acid ABC transporter, amino acid-binding protein
ORF01072	ribosomal protein S20 (rpsT)
ORF01073	pantothenate kinase (coaA)
ORF01074	conserved hypothetical protein
ORF01075	cytidine deaminase (cdd)
ORF01076	expressed putative lipoprotein
ORF01077	sugar ABC transporter, ATP-binding protein
ORF01078	sugar ABC transporter, permease protein, putative
ORF01079	sugar ABC transporter, permease protein, putative
ORF01080	NADH oxidase (nox-2)
ORF01081	L-lactate dehydrogenase (ldh)
ORF01082	DNA gyrase, A subunit (gyrA)
ORF01083	sortase SrtA (srtA)
ORF01089	GMP synthase (guaA)
ORF01090	transcriptional regulator, GntR family
ORF01091	gid protein (gid)
ORF01093	expressed putative lipoprotein
ORF01097	ABC transporter, ATP-binding protein
ORF01099	DNA-binding response regulator
ORF01101	site-specific recombinase, phage integrase family
ORF01106	signal recognition particle protein Ffh (ffh)
ORF01108	conserved hypothetical protein
ORF01109	sensor histidine kinase CiaH
ORF01110	DNA-binding response regulator CiaR (ciaR)
ORF01111	aminopeptidase N (pepN)

Tabl 8: GBS genes shared with GAS and pn um coccus

ORFxxxxx Ann tati n

ORF01112	phosphate transport system regulatory protein PhoU (phoU)
ORF01113	phosphate ABC transporter, ATP-binding protein PstB, putative
ORF01114	phosphate ABC transporter, ATP-binding protein PstB, putative
ORF01115	phosphate ABC transporter, permease protein PstA, putative
ORF01116	phosphate ABC transporter, permease protein
ORF01117	phosphate ABC transporter, phosphate-binding protein
ORF01118	NOL1/NOP2/sun family protein
ORF01119	inositol monophosphatase family protein
ORF01120	conserved hypothetical protein
ORF01121	conserved hypothetical protein
ORF01122	macrolide-efflux protein mreA/riboflavin biosynthesis protein RibF
ORF01123	tRNA pseudouridine synthase B (truB)
ORF01125	conserved hypothetical protein
ORF01128	permease, putative
ORF01129	ABC transporter, ATP-binding protein
ORF01131	DNA topoisomerase I (topA)
ORF01132	DprA/SMF protein, putative DNA processing factor (dprA)
ORF01134	iron compound ABC transporter, ATP-binding protein
ORF01137	acetyltransferase, CysE/LacA/LpxA/NodL family
ORF01138	ribonuclease HII (mhB)
ORF01139	GTP-binding protein
ORF01176	carbamoyl-phosphate synthase, large subunit (carB)
ORF01177	carbamoyl-phosphate synthase, small subunit (carA)
ORF01178	aspartate carbamoyltransferase (pyrB)
ORF01179	dihydroorotase, multifunctional complex type (pyrC)
ORF01180	orotate phosphoribosyltransferase (pyrE)
ORF01181	orotidine 5'-phosphate decarboxylase (pyrF)
ORF01183	ABC transporter, ATP-binding protein
ORF01184	ribonucleotide reductase, truncation
ORF01188	cardiolipin synthetase (cls)
ORF01189	formate-tetrahydrofolate ligase (fhs)
ORF01190	lipoate-protein ligase A (lplA)
ORF01198	flavoprotein-related protein
ORF01199	flavoprotein family protein
ORF01200	membrane protein, putative
ORF01201	phosphoglucosyltransferase (pgm)
ORF01203	IS861, transposase OrfB
ORF01205	ABC transporter, ATP-binding/permease protein
ORF01206	ABC transporter, ATP-binding/permease protein
ORF01207	conserved hypothetical protein
ORF01208	conserved hypothetical protein
ORF01209	Serine hydroxymethyltransferase
ORF01210	Sua5/YciO/YrdC/YwC family protein
ORF01211	modification methylase, HemK family
ORF01212	peptide chain release factor 1 (prfA)
ORF01213	thymidine kinases (tdk)
ORF01214	4-oxalocrotonate tautomerase (xylM)
ORF01216	ApbE family protein
ORF01220	xanthine permease (pbuX)
ORF01221	xanthine phosphoribosyltransferase (xpt)
ORF01222	guanosine monophosphate reductase (guaC)
ORF01227	phosphate acetyltransferase
ORF01228	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF01229	expressed protein of unknown function
ORF01230	GTP pyrophosphokinase family protein

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotati n

ORF01231 conserved hypothetical protein
ORF01232 ribose-phosphate pyrophosphokinase (prsA)
ORF01233 cysteine desulphurase (iscS)
ORF01234 conserved hypothetical protein
ORF01235 conserved hypothetical protein
ORF01236 DNA repair protein RadC (radC)
ORF01238 6-phospho-beta-glucosidase (ascB)
ORF01239 platelet activating factor, putative
ORF01240 hydrolase, haloacid dehalogenase-like family
ORF01242 voltage-gated chloride channel family protein
ORF01243 spermidine/putrescine ABC transporter, spermidine/putrescine-binding protein (potD)
ORF01244 spermidine/putrescine ABC transporter, permease protein (potC)
ORF01245 spermidine/putrescine ABC transporter, permease protein (potB)
ORF01246 spermidine/putrescine ABC transporter, ATP-binding protein (potA)
ORF01247 UDP-N-acetylenolpyruvoylglucosamine reductase (murB)
ORF01248 2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase (folK)
ORF01250 dihydropteroate synthase (folP)
ORF01251 GTP cyclohydrolase I (folE)
ORF01252 folylpolyglutamate synthase (folC)
ORF01259 aldehyde dehydrogenase family protein
ORF01260 membrane protein
ORF01274 gls24 protein, putative
ORF01276 gls24 protein, putative
ORF01279 conserved hypothetical protein
ORF01282 ATP-dependent DNA helicase PcrA (pcrA)
ORF01283 conserved hypothetical protein, FRAMESHIFT
ORF01284 uracil permease (uraA)
ORF01285 sodium:alanine symporter family protein
ORF01286 cation efflux family protein
ORF01290 ribosomal protein S1 (rpsA)
ORF01292 branched-chain amino acid aminotransferase (ilvE)
ORF01294 DNA topoisomerase IV, A subunit (parC)
ORF01295 DNA topoisomerase IV, B subunit (parE)
ORF01296 membrane protein, putative
ORF01297 uracil-DNA glycosylase (ung)
ORF01317 transcriptional regulator, LysR family, putative
ORF01319 purine nucleoside phosphorylase (deoD)
ORF01321 purine nucleoside phosphorylase (deoD)
ORF01323 phosphopentomutase (deoB)
ORF01324 ribose 5-phosphate isomerase (rpiA)
ORF01327 tributyrin esterase (estA)
ORF01328 metallo-beta-lactamase superfamily protein
ORF01329 ABC transporter, ATP-binding protein
ORF01330 ABC transporter, permease protein
ORF01331 conserved hypothetical protein
ORF01332 adherence and virulence protein A (pavA)
ORF01335 TPR domain protein
ORF01336 membrane protein
ORF01338 mutator MutT protein (mutX)
ORF01339 hyaluronidase
ORF01343 iminodiacetat oxidase, putative
ORF01344 conserved hypothetical protein TIGR00486
ORF01345 conserved hypothetical protein
ORF01346 DNA replication protein Dnad, putative
ORF01347 adenine phosphoribosyltransferas (apt)

Table 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF01350 single-stranded-DNA-specific exonuclease RecJ (recJ)
ORF01351 oxidoreductase, short chain dehydrogenase/reductase family
ORF01352 metallo-beta-lactamase superfamily protein
ORF01353 conserved hypothetical protein
ORF01354 GTP-binding protein HflX (hflX)
ORF01355 tRNA delta(2)-isopentenylpyrophosphate transferase (miaA)
ORF01357 exfoliative toxin A, putative
ORF01358 pullulanase, putative
ORF01362 conserved hypothetical protein
ORF01363 peptidase, M20/M25/M40 family
ORF01364 nitroreductase family protein
ORF01367 excinuclease ABC, C subunit (uvrC)
ORF01380 streptococcal histidine triad family protein
ORF01381 laminin-binding surface protein (lmb)
ORF01397 Tn5252, relaxase
ORF01403 mercuric reductase (merA)
ORF01406 IS861, transposase OrfB, truncation
ORF01407 cation-transporting ATPase, E1-E2 family
ORF01411 conserved hypothetical protein
ORF01412 cation-transporting ATPase, E1-E2 family
ORF01415 transcriptional repressor CopY, putative
ORF01416 cadmium resistance transporter, putative
ORF01451 C-5 cytosine-specific DNA methylase
ORF01453 conserved hypothetical protein
ORF01455 ribosomal protein L7/L12 (rplL)
ORF01456 ribosomal protein L10 (rplJ)
ORF01458 ATP-dependent Clp protease, ATP-binding subunit
ORF01467 GTP-binding protein (cgpA)
ORF01468 ATP-dependent Clp protease, ATP-binding subunit ClpX (clpX)
ORF01470 dihydrofolate reductase (folA)
ORF01471 thymidylate synthase (thyA)
ORF01472 HMG-CoA synthase
ORF01473 3-hydroxy-3-methylglutaryl-CoA reductase
ORF01474 conserved hypothetical protein
ORF01475 hemolysin III, putative
ORF01476 conserved hypothetical protein TIGR00147
ORF01479 isopentenyl-diphosphate delta-isomerase
ORF01480 phosphomevalonate kinase
ORF01481 diphosphomevalonate decarboxylase (mvaD)
ORF01482 mevalonate kinase, putative
ORF01484 DNA-binding response regulator
ORF01491 polypeptide deformylase, putative
ORF01495 ABC transporter, ATP-binding/permease protein
ORF01496 ABC transporter, ATP-binding/permease protein
ORF01498 ABC transporter, ATP-binding protein
ORF01499 polyA polymerase family protein
ORF01500 DegV family protein
ORF01501 expressed protein of unknown function
ORF01504 PTS system, fructose specific IIABC components
ORF01505 1-phosphofructokinase (fruK)
ORF01506 lactose phosphotransferase system repressor (lacR)
ORF01507 beta-lactam resistance factor
ORF01511 pyridine nucleotide-disulphide oxidoreductase family protein
ORF01512 tRNA (guanine-N1)-methyltransferase (trmD)
ORF01513 16S rRNA processing protein RimM (rimM)

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF01515	transcriptional regulator, RofA family
ORF01516	KH domain protein
ORF01517	ribosomal protein S16 (rpsP)
ORF01518	permease, putative
ORF01519	ABC transporter, ATP-binding protein
ORF01520	conserved hypothetical protein
ORF01523	carbamoyl-phosphate synthase, small subunit (carA)
ORF01524	pyrimidine operon regulatory protein (pyrR)
ORF01525	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF01526	lipoprotein signal peptidase (lspA)
ORF01527	transcriptional regulator, LysR family
ORF01528	ribosomal protein L27 (rpmA)
ORF01529	conserved hypothetical protein
ORF01530	ribosomal protein L21 (rplU)
ORF01531	conserved hypothetical protein, FRAMESHIFT
ORF01532	thiamine biosynthesis protein ThiI (thiI)
ORF01533	cysteine desulphurase (iscS)
ORF01536	glutathione reductase (gor)
ORF01537	conserved hypothetical protein
ORF01538	chorismate synthase (aroC)
ORF01539	3-dehydroquinate synthase (aroB)
ORF01540	3-dehydroquinate dehydratase (aroD)
ORF01541	conserved hypothetical protein
ORF01543	ribosomal protein L20 (rplT)
ORF01544	ribosomal protein L35 (rpmI)
ORF01545	translation initiation factor IF-3 (infC)
ORF01546	cytidylate kinase (cmk)
ORF01548	ferredoxin, 4Fe-4S
ORF01550	peptidase t (pepT)
ORF01551	polysaccharide biosynthesis protein, putative
ORF01552	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (murE)
ORF01553	iron compound ABC transporter, ATP-binding protein (fepC)
ORF01555	iron compound ABC transporter, permease protein
ORF01556	iron compound ABC transporter, permease protein
ORF01558	inorganic pyrophosphatase, manganese-dependent (ppa)
ORF01559	pyruvate formate-lyase-activating enzyme (pflA)
ORF01560	CBS domain protein
ORF01561	conserved hypothetical protein
ORF01564	PAP2 family protein
ORF01565	membrane protein, putative
ORF01567	expressed sortase family protein
ORF01568	sortase family protein
ORF01571	rogB protein FRAMESHIFT (rogB)
ORF01587	conserved hypothetical protein
ORF01589	RNA polymerase sigma-70 factor (rpoD)
ORF01590	DNA primase (dnaG)
ORF01591	large conductance mechanosensitive channel protein (mscL)
ORF01592	ribosomal protein S21 (rpsU)
ORF01594	amino acid ABC transporter, amino acid-binding protein
ORF01598	rhodanese family protein
ORF01602	glycogen phosphorylase (glgP)
ORF01603	4-alpha-glucanotransferase (malQ)
ORF01604	maltose operon repressor MalR, putative
ORF01605	maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein
ORF01606	maltose ABC transporter, permease protein

Tabl 8: GBS genes shar d with GAS and pneum c ccus

ORFxxxx Annotati n

ORF01607	maltose ABC transporter, permease protein
ORF01614	preprotein translocase SecA subunit, putative
ORF01619	preprotein translocase SecY family protein
ORF01634	excinuclease ABC, B subunit (uvrB)
ORF01636	glutamine ABC transporter, glutamine-binding protein/permease protein (glnP)
ORF01637	glutamine ABC transporter, ATP-binding protein, GlnQ putative
ORF01640	GTP-binding protein, GTP1/Obg family (obg)
ORF01646	amidase family protein
ORF01647	ribosomal small subunit pseudouridine synthase A (rsuA)
ORF01648	oxidoreductase, aldo/keto reductase family
ORF01651	lactoylglutathione lyase (gloA)
ORF01652	glycosyl transferase, group 2 family protein
ORF01654	SsrA-binding protein (smpB)
ORF01655	exoribonuclease, VacB/Rnb family (vacB)
ORF01657	preprotein translocase, SecG subunit
ORF01658	multi-drug resistance protein
ORF01662	dephospho-CoA kinase
ORF01663	formamidopyrimidine-DNA glycosylase (mutM)
ORF01677	GTP-binding protein Era (era)
ORF01678	diacylglycerol kinase (dgkA)
ORF01679	conserved hypothetical protein TIGR00043
ORF01685	PhoH family protein
ORF01687	conserved hypothetical protein
ORF01689	conserved hypothetical protein
ORF01690	ribosome recycling factor (frr)
ORF01691	uridylate kinase (pyrH)
ORF01693	peptide ABC transporter, ATP-binding protein FRAMESHIFT
ORF01697	ribosomal protein L1 (rplA)
ORF01698	ribosomal protein L11 (rplK)
ORF01706	IS861, transposase OrfB
ORF01707	chorismate binding enzyme
ORF01708	FtsK/SpolIIE family protein
ORF01709	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
ORF01710	manganese ABC transporter, permease protein
ORF01711	manganese ABC transporter, ATP-binding protein
ORF01712	manganese ABC transporter, manganese-binding adhesion liprotein
ORF01713	iron-dependent transcriptional regulator
ORF01714	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase (pfs)
ORF01716	MutT/nudix family protein
ORF01718	UDP-N-acetylglucosamine pyrophosphorylase (glmU)
ORF01722	oxidoreductase, Gfo/ldh/MocA family
ORF01725	gluconate 5-dehydrogenase, putative
ORF01726	conserved hypothetical protein
ORF01738	branched-chain amino acid transport system II carrier protein (brnQ)
ORF01739	methionyl-tRNA synthetase (metG)
ORF01745	exodeoxyribonuclease (exoA)
ORF01746	conserved hypothetical protein
ORF01752	copper homeostasis protein CutC, putative
ORF01755	tetrapyrrole methylase family protein
ORF01756	conserved hypothetical protein
ORF01758	DNA polymerase III, delta prime subunit, putative
ORF01759	thymidylate kinase (tmk)
ORF01773	ATP-dependent Clp protease, proteolytic subunit ClpP (clpP)
ORF01774	uracil phosphoribosyltransferase (upp)
ORF01777	RNA methyltransferase, TrmH family, group 2

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF01781 conserved hypothetical protein TIGR00278
ORF01782 ribosomal large subunit pseudouridine synthase B (rluB)
ORF01783 conserved hypothetical protein TIGR00281
ORF01784 conserved hypothetical protein
ORF01785 integrase/recombinase, phage integrase family
ORF01786 CBS domain protein
ORF01787 conserved hypothetical protein
ORF01788 HAM1 protein
ORF01789 glutamate racemase (muri)
ORF01791 membrane protein, putative
ORF01792 transcriptional regulator, biotin repressor family
ORF01793 membrane protein, putative
ORF01795 RNA methyltransferase, TrmH family
ORF01796 acylphosphatase
ORF01797 lipoprotein, putative
ORF01799 amino acid ABC transporter, permease protein
ORF01801 amidase family protein
ORF01802 transcription elongation factor GreA (greA)
ORF01803 conserved hypothetical protein
ORF01804 acetyltransferase, GNAT family
ORF01805 UDP-N-acetylmuramate--alanine ligase (murC)
ORF01806 conserved hypothetical protein
ORF01808 expressed putative helicase
ORF01811 phosphoglycerate dehydrogenase-related protein
ORF01812 primosomal protein DnaI (dnaI)
ORF01813 conserved hypothetical protein
ORF01814 conserved hypothetical protein TIGR00244
ORF01815 sensor histidine kinase CsrS (csrS)
ORF01816 DNA-binding response regulator CsrR (csrR)
ORF01817 conserved hypothetical protein
ORF01818 heat shock protein HtpX (htpX)
ORF01820 lemA protein (lemA)
ORF01821 glucose-inhibited division protein B (gidB)
ORF01822 sodium transport family protein
ORF01823 potassium uptake protein, Trk family, putative
ORF01825 ABC transporter, ATP-binding protein
ORF01828 branched-chain amino acid transport system II carrier protein (brnQ)
ORF01829 alcohol dehydrogenase, zinc-containing (adh)
ORF01830 ABC transporter, permease protein
ORF01831 ABC transporter, ATP-binding protein
ORF01833 expressed YaeC family protein
ORF01834 ABC transporter, substrate-binding protein
ORF01835 glutamine amidotransferase, class I
ORF01837 conserved hypothetical protein TIGR01033
ORF01846 glycerol uptake facilitator protein (glpF)
ORF01849 conserved hypothetical protein
ORF01851 conserved hypothetical protein
ORF01852 lojap-related protein
ORF01854 conserved hypothetical protein TIGR00488
ORF01855 conserved hypothetical protein TIGR00482
ORF01856 conserved hypothetical protein TIGR00253
ORF01857 GTP-binding protein
ORF01858 hydrolase, haloacid dehalogenase-like family
ORF01860 glutamyl-tRNA(Gln) amidotransferase, B subunit (gatB)
ORF01861 glutamyl-tRNA(Gln) amidotransferase, A subunit (gatA)

Tabl 8: GBS genes shared with GAS and pneumoc ccus

ORFxxxxx Annotati n

ORF01862	glutamyl-tRNA(Gln) amidotransferase, C subunit (gatC)
ORF01867	isochorismatase family protein
ORF01869	transcriptional regulator CodY, putative
ORF01870	aminotransferase, class I
ORF01871	universal stress protein family FRAMESHIFT
ORF01872	hydrolase, haloacid dehalogenase-like family
ORF01873	asparaginase family protein
ORF01874	shikimate 5-dehydrogenase (aroE)
ORF01876	ATP-dependent DNA helicase RecG (recG)
ORF01878	alanine racemase (alr)
ORF01879	holo-(acyl-carrier-protein) synthase (acpS)
ORF01881	preprotein translocase, SecA subunit (secA)
ORF01882	mannose-6-phosphate isomerase, class I (manA)
ORF01883	fructokinase (scrK)
ORF01885	PTS system IIBC components
ORF01886	sucrose-6-phosphate hydrolase (scrB)
ORF01887	sucrose operon repressor ScrR (scrR)
ORF01888	N utilization substance protein B (nusB)
ORF01889	conserved hypothetical protein
ORF01890	translation elongation factor P (efp)
ORF01900	cytidine/deoxycytidylate deaminase family protein
ORF01906	excinuclease ABC, A subunit (uvrA)
ORF01907	conserved hypothetical protein
ORF01908	magnesium transporter, CorA family (corA)
ORF01909	ribosomal protein S18 (rpsR)
ORF01910	single-strand binding protein (ssb)
ORF01911	ribosomal protein S6 (rpsF)
ORF01912	A/G-specific adenine glycosylase (mutY)
ORF01914	thioredoxin (trx)
ORF01915	PAP2 family protein
ORF01916	MutS2 family protein
ORF01917	conserved hypothetical protein
ORF01918	conserved hypothetical protein
ORF01919	ribonuclease HIII (rnhC)
ORF01920	signal peptidase I
ORF01921	helicase, putative
ORF01923	DNA-damage inducible protein P (dinP)
ORF01924	formate acetyltransferase (pflD)
ORF01926	conserved hypothetical protein
ORF01927	proteinase, putative, degenerate, FRAMESHIFT
ORF01929	glycerol uptake facilitator protein, putative
ORF01930	universal stress protein family
ORF01933	X-pro dipeptidyl-peptidase (pepX)
ORF01937	ABC transporter, ATP-binding protein CydC (cydC)
ORF01938	ABC transporter, ATP-binding protein CydD
ORF01945	conserved hypothetical protein TIGR00103
ORF01948	exonuclease
ORF01949	conserved hypothetical protein
ORF01950	conserved hypothetical protein TIGR00275
ORF01952	ribosomal protein S14 (rpsN)
ORF01957	O-sialoglycoprotein endopeptidase family protein
ORF01958	ribosomal-protein-alanine acetyltransferase, putative
ORF01960	expressed protein of unknown function
ORF01961	conserved hypothetical protein
ORF01962	metallo-beta-lactamase superfamily protein

Table 8: GBS genes shared with GAS and pn umoc ccus

ORFxxxxx Annotation

ORF01963 conserved hypothetical protein
ORF01964 glutamine synthetase, type I (glnA)
ORF01965 transcriptional regulator GlnR (glnR)
ORF01967 conserved hypothetical protein
ORF01969 phosphoglycerate kinase (pgk)
ORF01971 glyceraldehyde 3-phosphate dehydrogenase (gap)
ORF01972 translation elongation factor G (fusA)
ORF01973 ribosomal protein S7 (rpsG)
ORF01974 ribosomal protein S12 (rpsL)
ORF01975 pur operon repressor (purR)
ORF01976 HD domain protein
ORF01977 conserved hypothetical protein
ORF01978 conserved hypothetical protein
ORF01979 ribulose-phosphate 3-epimerase (rpe)
ORF01980 conserved hypothetical protein TIGR00157
ORF01983 dimethyladenosine transferase (ksgA)
ORF01985 primase-related protein
ORF01987 deoxyribonuclease, TatD family
ORF01992 dltD protein (dltD)
ORF01993 D-alanyl carrier protein (dltC)
ORF01994 dltB protein (dltB)
ORF01996 D-alanine-activating enzyme (dltA)
ORF01997 sensor histidine kinase
ORF01998 DNA-binding response regulator
ORF01999 ribosomal protein L34 (rpmH)
ORF02004 amino acid ABC transporter, ATP-binding protein
ORF02007 conserved hypothetical protein
ORF02008 transcriptional antiterminator, BglG family
ORF02017 sugar binding transcriptional regulator, LacI family
ORF02018 transaldolase family protein
ORF02019 carbohydrate isomerase, AraD/FucA family
ORF02020 hexulose-6-phosphate isomerase, putative
ORF02021 hexulose-6-phosphate synthase, putative
ORF02022 PTS system, IIA component
ORF02023 PTS system, IIB component
ORF02024 transport protein SgaT, putative
ORF02027 adenylosuccinate synthetase (purA)
ORF02033 chaperonin, 33 kDa (hslO)
ORF02034 NifR3/Smm1 family protein
ORF02037 ATP-dependent Clp protease, ATP-binding subunit
ORF02038 transcriptional regulator CtsR (ctsR)
ORF02040 translation elongation factor Ts (tsf)
ORF02041 ribosomal protein S2 (rpsB)
ORF02043 alkyl hydroperoxide reductase, subunit F (ahpF)
ORF02076 prophage LambdaSa2, single-strand binding protein (ssb)
ORF02082 prophage LambdaSa2, type II DNA modification methyltransferase, putative
ORF02086 prophage LambdaSa2, replicative DNA helicase (dnaC)
ORF02104 endopeptidase O (pepO)
ORF02110 polypeptide deformylase (def)
ORF02111 sugar binding transcriptional regulator RegR (regR)
ORF02112 conserved hypothetical protein
ORF02113 PTS system, IID component
ORF02114 PTS system, IIC component
ORF02115 PTS system, IIB component
ORF02116 glucuronyl hydrolase

Tabl 8: GBS g nes har d with GAS and pn umococcus

ORFxxxxx Ann tation

ORF02118	PTS system, IIA component
ORF02120	oxidoreductase, short-chain dehydrogenase/reductase family
ORF02121	conserved hypothetical protein
ORF02122	carbohydrate kinase, PfkB family
ORF02123	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase (eda)
ORF02127	DNA polymerase III, alpha subunit, Gram-positive type
ORF02129	prolyl-tRNA synthetase (proS)
ORF02130	membrane-associated zinc metalloprotease, putative
ORF02131	phosphatidate cytidyltransferase (cdsA)
ORF02132	undecaprenyl diphosphate synthase (uppS)
ORF02133	preprotein translocase, YajC subunit (yajC)
ORF02140	glucan 1,6-alpha-glucosidase (dexB)
ORF02141	sugar ABC transporter, ATP-binding protein (msmK)
ORF02142	helix-turn-helix domain protein, fis-type
ORF02144	tagatose 1,6-diphosphate aldolase (lacD)
ORF02145	tagatose-6-phosphate kinase (lacC)
ORF02146	galactose-6-phosphate isomerase, LacB subunit (lacB)
ORF02147	galactose-6-phosphate isomerase, LacA subunit (lacA)
ORF02149	PTS system, IIC component, putative
ORF02150	PTS system, IIB component, putative
ORF02152	PTS system, IIA component, putative
ORF02153	lactose phosphotransferase system repressor (lacR)
ORF02157	adhesion lipoprotein
ORF02158	expressed protein of unknown function TIGR00256
ORF02159	GTP pyrophosphokinase (relA)
ORF02161	nrdI protein (nrdI)
ORF02164	iron ABC transporter, iron-binding protein
ORF02165	DNA-binding response regulator
ORF02167	PTS system, IID component
ORF02168	PTS system, IIC component
ORF02174	ABC transporter, ATP-binding protein
ORF02176	response regulator
ORF02177	conserved hypothetical protein
ORF02178	PTS system, IIABC components
ORF02179	sensor histidine kinase
ORF02180	phosphate regulon response regulator PhoB (phoB)
ORF02182	phosphate ABC transporter, ATP-binding protein (pstB)
ORF02183	phosphate ABC transporter, permease protein
ORF02184	phosphate ABC transporter, permease protein
ORF02188	conserved hypothetical protein TIGR00046
ORF02189	ribosomal protein L11 methyltransferase (prmA)
ORF02197	conserved hypothetical protein
ORF02199	ATPase, AAA family
ORF02249	mercuric reductase (merA)
ORF02272	DNA topology modulation protein FlaR, putative
ORF02273	glycerol dehydrogenase, putative
ORF02281	DNA-binding response regulator
ORF02285	leucyl-tRNA synthetase (leuS)
ORF02290	transcription antitermination protein NusG (nusG)
ORF02293	penicillin-binding protein 2A (pbp2A)
ORF02294	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF02296	phosphopentomutase (deoB)
ORF02297	deoxyribose-phosphate aldolase (deoC)
ORF02300	uridine phosphorylase (udp)
ORF02302	60 kda chaperonin (groEL)

Tabl 8: GBS genes shared with GAS and pneumococcus

ORFxxxxx Annotation

ORF02303	chaperonin, 10 kDa (groES)
ORF02305	ABC transporter, ATP-binding protein
ORF02306	ABC transporter, permease protein
ORF02307	expressed putative lipoprotein
ORF02309	glyoxalase family protein
ORF02310	conserved hypothetical protein
ORF02311	anaerobic ribonucleoside-triphosphate reductase activating protein (nrdG)
ORF02312	acetyltransferase, GNAT family
ORF02315	anaerobic ribonucleoside-triphosphate reductase (nrdD)
ORF02318	conserved hypothetical protein
ORF02320	conserved hypothetical protein
ORF02321	conserved hypothetical protein
ORF02322	recA protein (recA)
ORF02325	DNA-3-methyladenine glycosylase I (tag)
ORF02327	Holliday junction DNA helicase RuvA (ruvA)
ORF02329	DNA mismatch repair protein HexB (hexB)
ORF02333	arginine repressor ArgR, putative
ORF02334	arginyl-tRNA synthetase (argS)
ORF02337	conserved hypothetical protein
ORF02338	conserved hypothetical protein
ORF02339	aspartyl-tRNA synthetase (aspS)
ORF02340	histidyl-tRNA synthetase (hisS)
ORF02342	ribosomal protein L33 (rpmG)
ORF02357	DNA-binding response regulator
ORF02359	membrane protein, putative
ORF02360	carbamate kinase (arcC)
ORF02361	ornithine carbamoyltransferase (argF)
ORF02364	amino acid ABC transporter, ATP-binding protein
ORF02365	amino acid ABC transporter, permease and amino acid-binding protein
ORF02370	membrane protein, putative
ORF02371	transcriptional regulator, TetR family, putative
ORF02373	ribosomal protein S4 (rpsD)
ORF02374	conserved hypothetical protein
ORF02375	replicative DNA helicase (dnaC)
ORF02376	ribosomal protein L9 (rplI)
ORF02377	DHH family protein
ORF02378	glucose inhibited division protein A (gidA)
ORF02380	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (trmU)
ORF02381	L-serine dehydratase, iron-sulfur-dependent, beta subunit (sdhB)
ORF02382	L-serine dehydratase, iron-sulfur-dependent, alpha subunit (sdhA)
ORF02385	cobalt transport family protein
ORF02386	ABC transporter, ATP-binding protein
ORF02387	ABC transporter, ATP-binding protein, FRAMESHIFT
ORF02388	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (pgsA)
ORF02389	peptidase, M16 family
ORF02390	conserved hypothetical protein
ORF02391	conserved hypothetical protein
ORF02392	recF protein (recF)
ORF02396	inosine-5'-monophosphate dehydrogenase (guaB)
ORF02397	transcriptional regulator, ArgR family
ORF02400	arginine deiminase (arcA)
ORF02402	ornithine carbamoyltransferase (argF)
ORF02404	carbamate kinase (arcC)
ORF02405	tryptophanyl-tRNA synthetase (trpS)
ORF02407	conserved hypothetical protein

Tabl 8: GBS genes shared with GAS and pn umoc ccus

ORFxxxxx Annotation

ORF02408	ABC transporter, ATP-binding protein
ORF02409	ABC transporter, permease protein, putative
ORF02410	conserved hypothetical protein TIGR00246
ORF02411	serine protease
ORF02412	partitioning protein, ParB family
ORF02413	chromosomal replication initiator protein DnaA (dnaA)
ORF02415	DNA polymerase III, beta subunit (dnaN)
ORF02417	conserved hypothetical protein
ORF02419	conserved hypothetical GTP-binding protein
ORF02420	peptidyl-tRNA hydrolase (pth)
ORF02421	transcription-repair coupling factor (mfd)
ORF02423	S4 domain protein
ORF02424	cell division protein DivIC, putative
ORF02426	expressed protein of unknown function
ORF02427	MesJ/Ycf62 family protein
ORF02429	cell division protein FtsH (ftsH)

Table 9: GBS genes shared with *pn umoccc cus*

ORFxxxxx Annotation

ORF00017	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase (purH)
ORF00025	conserved hypothetical protein
ORF00029	acetyl xylan esterase, putative
ORF00042	aldehyde-alcohol dehydrogenase (adhE)
ORF00044	threonine synthase (thrC)
ORF00081	ribosomal protein L17 (rplQ)
ORF00090	conserved hypothetical protein
ORF00129	argininosuccinate synthase (argG)
ORF00156	oligopeptide ABC transporter, substrate-binding protein, putative
ORF00189	protease, putative
ORF00194	thioredoxin family protein
ORF00195	tRNA binding domain protein
ORF00217	conserved domain protein
ORF00218	PTS system, IIB component, putative
ORF00220	transketolase, N-terminal subunit
ORF00221	transketolase, C-terminal subunit
ORF00223	oxidoreductase, putative
ORF00282	acetyltransferase, GNAT family
ORF00290	IS1381, transposase OrfB
ORF00291	IS1381, transposase OrfA
ORF00293	conserved hypothetical protein
ORF00301	membrane protein, putative
ORF00343	ABC transporter, permease protein, putative
ORF00344	conserved hypothetical protein
ORF00382	aspartate kinase family protein
ORF00399	conserved hypothetical protein
ORF00439	cell wall surface anchor family protein
ORF00447	cytidine/deoxycytidylate deaminase family protein
ORF00450	5-formyltetrahydrofolate cyclo-ligase family protein
ORF00480	transcriptional regulator, MerR family
ORF00499	acetyltransferase, GNAT family
ORF00504	magnesium transporter, CorA family
ORF00521	VanZF domain protein
ORF00612	IS1381, transposase OrfA
ORF00613	IS1381, transposase OrfB
ORF00690	transmembrane protein Vexp1 (vex1)
ORF00691	ABC transporter, ATP-binding protein Vexp2 (vex2)
ORF00692	transmembrane protein Vexp3 (vex3)
ORF00714	conserved hypothetical protein
ORF00732	expressed cell wall surface anchor family protein, putative
ORF00774	ABC transporter, ATP-binding protein
ORF00778	ABC transporter, ATP-binding protein
ORF00780	conserved hypothetical protein
ORF00790	beta-glucuronidase
ORF00800	alpha amylase family protein
ORF00807	amino acid ABC transporter, permease protein
ORF00809	amino acid ABC transporter, amino acid-binding protein
ORF00814	conserved hypothetical protein
ORF00823	bacterial luciferase family protein
ORF00840	riboflavin biosynthesis protein RibD (ribD)
ORF00841	riboflavin synthase, alpha subunit (ribE)
ORF00842	riboflavin biosynthesis protein RibA (ribA)
ORF00843	riboflavin synthase, beta subunit (ribH)
ORF00866	penicillin-binding protein 2b
ORF00905	membrane protein, putative

Table 9: GBS genes shared with pneumococcus

ORFxxxxx Annotation

ORF00910	major facilitator family protein
ORF00913	hydrolase, haloacid dehalogenase-like family
ORF00918	conserved hypothetical protein
ORF00945	conserved hypothetical protein
ORF00948	ABC transporter, ATP-binding protein
ORF00952	phosphomethylpyrimidine kinase (thiD)
ORF00953	hydroxyethylthiazole kinase (thiM)
ORF00954	thiamine-phosphate pyrophosphorylase (thiE)
ORF00961	GtrA family protein
ORF00967	1,4-alpha-glucan branching enzyme (glgB)
ORF00968	glucose-1-phosphate adenylyltransferase (glgC)
ORF00971	glycogen synthase (glgA)
ORF00985	acetyltransferase, GNAT family
ORF00990	magnesium transporter, CorA family, putative
ORF01022	nucleoside diphosphate kinase (ndk)
ORF01031	nucleoside diphosphate kinase domain protein
ORF01085	conserved hypothetical protein
ORF01087	IS1381, transposase OrfA
ORF01088	IS1381, transposase OrfB
ORF01098	ABC transporter, permease protein, putative
ORF01100	sensor histidine kinase
ORF01102	ABC transporter, substrate-binding protein
ORF01127	protease, putative
ORF01135	iron compound ABC transporter, permease protein
ORF01136	iron compound ABC transporter, permease protein
ORF01185	aspartate-semialdehyde dehydrogenase (asd)
ORF01217	conserved hypothetical protein
ORF01218	conserved hypothetical protein
ORF01219	formate/nitrite transporter family protein
ORF01226	oxidoreductase, short chain dehydrogenase/reductase family, FRAMESHIFT
ORF01254	homoserine kinase (thrB)
ORF01255	homoserine dehydrogenase (hom)
ORF01264	transcriptional regulator, Cro/C1 family
ORF01268	thiol peroxidase (psaD)
ORF01305	glycosyltransferase CpsJ(V) (cpsJ)
ORF01306	glycosyltransferase CpsO(V) (cpsO)
ORF01313	CpsD protein (cpsD)
ORF01314	cpsC protein (cpsC)
ORF01315	capsular polysaccharide biosynthesis protein CpsB (cpsB)
ORF01316	capsular polysaccharide biosynthesis protein CpsA (cpsA)
ORF01326	conserved hypothetical protein
ORF01333	alpha-acetolactate decarboxylase (budA)
ORF01334	acetolactate synthase, catabolic (ilvK)
ORF01337	MutT/nudix family protein
ORF01369	MATE efflux family protein
ORF01398	Tn5252, Orf 9 protein
ORF01399	Tn5252, Orf 10 protein
ORF01446	protease, putative
ORF01447	conserved hypothetical protein
ORF01449	conserved hypothetical protein
ORF01492	NADP-specific glutamate dehydrogenase (gdhA)
ORF01569	expressed cell wall surface anchor family protein
ORF01570	cell wall surface anchor family protein
ORF01574	polysaccharide biosynthesis protein
ORF01579	nucleotidyl transferase, putative

Table 9: GBS genes shared with pneumococcus

ORFxxxxx Annotation

ORF01580 polysaccharide biosynthesis protein, putative
ORF01612 conserved hypothetical protein
ORF01613 glycosyl transferase, group 1 family protein
ORF01617 conserved hypothetical protein
ORF01618 conserved hypothetical protein
ORF01621 glycosyl transferase, putative
ORF01622 glycosyl transferase, group 2 family protein
ORF01623 glycosyl transferase, family 8, degenerate
ORF01624 IS1381, transposase OrfB
ORF01625 IS1381, transposase OrfA
ORF01626 glycosyl transferase family 8
ORF01627 glycosyl transferase, family 8
ORF01628 conserved hypothetical protein
ORF01630 cell wall surface anchor family protein
ORF01635 protease, putative
ORF01643 aminopeptidase PepS (pepS)
ORF01702 peptidase, M20/M25/M40 family
ORF01731 IS1381, transposase OrfA
ORF01732 IS1381, transposase OrfB
ORF01740 tellurite resistance protein TehB (tehB)
ORF01747 methylated-DNA--protein-cysteine S-methyltransferase (ogt)
ORF01749 acetyltransferase, GNAT family
ORF01763 AcuB family protein
ORF01764 branched-chain amino acid ABC transporter, ATP-binding protein (livF)
ORF01765 branched-chain amino acid ABC transporter, ATP-binding protein (livG)
ORF01766 branched-chain amino acid ABC transporter, permease protein
ORF01767 branched-chain amino acid ABC transporter, permease protein (livH)
ORF01769 branched-chain amino acid ABC transporter, amino acid-binding protein
ORF01775 aminotransferase, class I
ORF01779 potassium uptake protein, Trk family
ORF01780 cation uptake protein, Trk family
ORF01824 cobalt transport family protein
ORF01826 conserved hypothetical protein
ORF01832 peptidase, M20/M25/M40 family
ORF01845 conserved hypothetical protein
ORF01848 transcriptional regulator, MerR family
ORF01853 isochorismatase family protein
ORF01859 membrane protein
ORF01875 oxidoreductase, aldo/keto reductase family
ORF01880 phospho-2-dehydro-3-deoxyheptonate aldolase
ORF01981 rRNA (guanine-N1-)-methyltransferase, putative
ORF02083 prophage LambdaSa2, DNA replication protein DnaC, putative
ORF02101 Na ⁺ /H ⁺ exchanger family protein
ORF02107 membrane protein, putative
ORF02139 UDP-glucose 4-epimerase (galE)
ORF02143 lacX protein
ORF02162 conserved hypothetical protein
ORF02186 hemolysin precursor, putative
ORF02192 transcriptional regulator, MerR family
ORF02195 MutT/nudix family protein
ORF02228 IS1381, transposase OrfB
ORF02229 IS1381, transposase OrfA
ORF02233 conserved hypothetical protein
ORF02234 conserved hypothetical protein
ORF02276 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (metE)

2043517 1537752

Tabl 9: GBS g n shar d with pn umoccocus

ORFxxxxx Annotati n

ORF02278	branched-chain amino acid transport protein AziC, putative
ORF02288	glycosyl transferas , family 8
ORF02289	glycosyl transferase, family 8
ORF02341	ribosomal protein L32 (rpmF)
ORF02343	conserved hypothetical protein
ORF02358	sensor histidine kinase
ORF02369	conserved hypothetical protein
ORF02384	LysM domain protein
ORF02428	hypoxanthine-guanine phosphoribosyltransferase (hpt)
ORF03011	ribosomal protein L33
ORF03014	ribosomal protein L33

Table 10: GBS genes shared with GAS

ORFxxxxx Annotation

ORF00064	ribosomal protein S14, putative
ORF00095	D-alanyl-D-alanine carboxypeptidase family protein
ORF00096	N-acetylmuramoyl-L-alanine amidase, family 4 protein
ORF00110	conserved hypothetical protein
ORF00112	DNA repair protein RadA (radA)
ORF00124	permease, putative
ORF00148	glycosyl transferase, group 4 family protein
ORF00154	penicillin-binding protein 4, putative
ORF00157	oligopeptide ABC transporter, permease protein
ORF00206	oligopeptide ABC transporter, oligopeptide-binding protein
ORF00207	oligopeptide ABC transporter, permease protein
ORF00208	oligopeptide ABC transporter, permease protein
ORF00209	peptide ABC transporter, ATP-binding protein
ORF00210	peptide ABC transporter, ATP-binding protein
ORF00216	IS1548, transposase
ORF00226	conserved hypothetical protein
ORF00232	conserved hypothetical protein
ORF00239	site-specific recombinase, phage integrase family
ORF00250	conserved hypothetical protein
ORF00251	conserved hypothetical protein
ORF00289	ABC transporter, ATP-binding protein
ORF00305	NADH oxidase, putative
ORF00317	cell division protein FtsL, putative
ORF00333	conserved hypothetical protein
ORF00383	hydrolase, haloacid dehalogenase-like family
ORF00430	expressed putative lipoprotein
ORF00431	transcriptional repressor CopY
ORF00434	membrane protein, putative
ORF00438	transcriptional regulator, Fur family
ORF00442	membrane protein, putative
ORF00445	bioY family protein
ORF00446	AtsA/ElaC family protein
ORF00468	expressed putative protease
ORF00469	glycosyl transferase, group 2 family protein
ORF00471	nrpI protein (nrpI)
ORF00473	expressed protein of unknown function
ORF00474	conserved hypothetical protein
ORF00507	conserved hypothetical protein
ORF00525	bioY family protein
ORF00528	thiolase
ORF00531	AMP-binding enzyme domain protein
ORF00548	YGGT family protein
ORF00565	exodeoxyribonuclease VII, small subunit (xseB)
ORF00568	arginine repressor ArgR, putative
ORF00572	expressed putative lipase/acylhydrolase
ORF00573	conserved hypothetical protein
ORF00586	iron-sulfur cluster-binding protein, putative
ORF00592	oxidoreductase, short chain dehydrogenase/reductase family
ORF00604	dipeptidase
ORF00611	voltage-gated chloride channel family protein
ORF00619	prophage LambdaSa1, repressor protein, putative
ORF00622	conserved hypothetical protein
ORF00627	prophage LambdaSa1, antirepressor, putative
ORF00634	conserved hypothetical protein
ORF00648	conserved hypothetical protein

Tabl 10: GBS g n s shar d with GAS

ORFxxxxx Annotation

ORF00654 conserved hypothetical prot in
ORF00655 conserved hypothetical protein
ORF00656 conserved hypothetical protein
ORF00658 conserved hypothetical protein
ORF00659 conserved hypothetical protein
ORF00660 prophage LambdaSa1, structural protein, putative
ORF00662 conserved hypothetical protein
ORF00663 conserved hypothetical protein
ORF00664 conserved hypothetical protein
ORF00665 conserved hypothetical protein
ORF00666 prophage LambdaSa1, structural protein
ORF00668 conserved hypothetical protein
ORF00669 prophage LambdaSa1, pblA protein, internal deletion
ORF00677 prophage LambdaSa1, lysin, putative
ORF00679 conserved hypothetical protein
ORF00695 transposase OrfB, IS3 family, truncation
ORF00697 conserved hypothetical protein
ORF00707 conserved domain protein
ORF00713 acid phosphatase precursor, class B
ORF00720 transposase OrfB, IS3 family FRAMESHIFT
ORF00721 transposase OrfA, IS3 family
ORF00751 cyaA protein (cyaA)
ORF00755 cyll protein (cyll)
ORF00760 serine protease, subtilase family, putative POINT MUTATION
ORF00781 transcriptional regulator, LysR family
ORF00783 regulatory protein, putative
ORF00785 IS1548, transposase
ORF00786 regulatory protein, putative, truncation
ORF00787 D-lactate dehydrogenase (ldhA)
ORF00801 glycosyl transferase, group 1 family protein
ORF00805 conserved hypothetical protein
ORF00826 phage shock protein C, putative
ORF00833 conserved hypothetical protein
ORF00845 hydrolase, haloacid dehalogenase-like family
ORF00852 conserved hypothetical protein
ORF00853 expressed putative lipoprotein
ORF00857 IS1548, transposase
ORF00890 conserved hypothetical protein
ORF00902 conserved hypothetical protein
ORF00926 membrane protein, putative
ORF00927 membrane protein, putative
ORF00987 conserved hypothetical protein
ORF01009 expressed protein of unknown function
ORF01010 lipoyl-binding domain protein
ORF01011 oxidoreductase, putative
ORF01012 conserved hypothetical protein
ORF01024 expressed putative lipoprotein
ORF01061 signal peptidase I, putative
ORF01064 IS1548, transposase
ORF01084 glyoxylase family protein
ORF01104 SatD
ORF01126 conserved hypothetical protein
ORF01191 conserved hypothetical protein
ORF01192 conserved hypothetical protein
ORF01193 glycine cleavage system H protein, putative

Tabl 10: GBS genes shar d with GAS

ORFxxxxx Ann tation

ORF01194 bacterial luciferas family protein
ORF01195 oxidoreductase, FMN-binding
ORF01197 lipoate-protein ligase A family protein
ORF01202 IS861, transposase OrfA
ORF01223 drug resistance transporter, EmrB/QacA family, putative
ORF01224 conserved hypothetical protein
ORF01225 potassium uptake protein, putative
ORF01237 membrane protein, putative
ORF01249 dihydroneopterin aldolase (folB)
ORF01256 polysaccharide deacetylase family protein
ORF01273 transcriptional regulator, GntR family/potassioum uptake protein, TrkA family
ORF01280 conserved hypothetical protein
ORF01281 conserved hypothetical protein
ORF01289 lipoprotein, putative
ORF01291 conserved hypothetical protein
ORF01298 conserved hypothetical protein
ORF01318 conserved hypothetical protein
ORF01320 voltage-gated chloride channel family protein, putative
ORF01322 arsenate reductase (arsC)
ORF01340 dTDP-glucose 4,6-dehydratase (rfbB)
ORF01341 dTDP-4-dehydrorhamnose 3,5-epimerase
ORF01342 glucose-1-phosphate thymidyltransferase (rfbA)
ORF01356 hypothetical protein
ORF01368 conserved hypothetical protein
ORF01374 ISSdy1, transposase OrfB
ORF01388 transposase OrfA, IS3 family
ORF01389 transposase OrfB, IS3 family, truncation
ORF01391 ISSdy1, transposase OrfB FRAMESHIFT
ORF01396 transcriptional regulator, Cro/Ci family
ORF01419 repressor protein, putative
ORF01461 amino acid permease
ORF01469 conserved hypothetical protein
ORF01483 sensor histidine kinase
ORF01485 GTP pyrophosphokinase family protein
ORF01490 5'-nucleotidase family protein
ORF01509 2-dehydropantoate 2-reductase, putative
ORF01510 regulatory protein, putative
ORF01522 carbamoyl-phosphate synthase, large subunit, putative
ORF01542 sulfatase
ORF01549 conserved hypothetical protein
ORF01554 iron compound ABC transporter, substrate-binding protein
ORF01557 conserved hypothetical protein
ORF01563 conserved hypothetical protein TIGR01212
ORF01583 glycosyltransferase, group 2 family protein
ORF01584 glycosyltransferase, group 2 family protein
ORF01585 glycosyltransferase, putative
ORF01586 dTDP-4-dehydrorhamnose reductase (rfbD)
ORF01593 conserved hypothetical protein
ORF01599 conserved hypothetical protein
ORF01600 glycerol-3-phosphate transporter, putative
ORF01639 conserved hypothetical protein
ORF01650 nitroreductase family protein
ORF01653 amino acid permease
ORF01665 transcriptional regulator, MutR family
ORF01683 MutT/nudix family protein

Table 10: GBS genes shared with GAS

ORFxxxxx Annotati n

ORF01686	67 kDa Myosin-crossreactive streptococcal antigen
ORF01688	peptid methionine sulfoxide reductase (msrA)
ORF01694	peptide ABC transporter, permease protein
ORF01704	conserved hypothetical protein
ORF01705	IS861, transposase OrfA
ORF01741	membrane protein, putative
ORF01770	conserved hypothetical protein
ORF01772	IS1548, transposase
ORF01790	conserved hypothetical protein
ORF01794	conserved hypothetical protein
ORF01800	amino acid ABC transporter, substrate-binding protein
ORF01810	IS1548, transposase
ORF01827	sodium:dicarboxylate symporter family protein
ORF01877	immunogenic secreted protein, putative
ORF01913	transcriptional regulator, Cro/C1 family
ORF01928	membrane protein, putative
ORF01931	transporter, putative
ORF01932	transcriptional regulator, Crp/Fnr family
ORF01947	transcriptional regulator, merR family
ORF01970	acid phosphatase
ORF02002	amino acid ABC transporter, permease protein
ORF02028	perfringolysin O regulator protein (pfoR)
ORF02029	conserved hypothetical protein
ORF02031	expressed protein of unknown function
ORF02032	expressed protein of unknown function
ORF02035	deoxynucleoside kinase family protein
ORF02042	alkyl hydroperoxide reductase, subunit C (ahpC)
ORF02126	transcriptional regulator, MarR family
ORF02128	N-acetylmuramoyl-L-alanine amidase, family 4 protein
ORF02135	malate oxidoreductase
ORF02136	citrate carrier protein, CCS family
ORF02137	sensor histidine kinase family protein
ORF02138	response regulator
ORF02166	conserved hypothetical protein
ORF02169	PTS system, IIB component
ORF02170	PTS system, IIA component, putative
ORF02202	ABC transporter, ATP-binding protein
ORF02262	ABC transporter, ATP-binding protein
ORF02270	cAMP factor (cfb)
ORF02280	serine protease, subtilase family, putative
ORF02286	major facilitator family protein
ORF02292	preprotein translocase, SecE subunit, putative
ORF02295	Lyme disease proteins of unknown function, putative
ORF02298	Na ⁺ dependent nucleoside transporter
ORF02301	transcriptional regulator, GntR family
ORF02313	virulence factor MviM, putative
ORF02316	membrane protein, putative
ORF02319	conserved hypothetical protein TIGR00250
ORF02328	transporter, putative
ORF02331	cold shock protein, CSD family
ORF02332	DNA mismatch repair protein HexA (hexA)
ORF02335	conserved hypothetical protein
ORF02372	conserved hypothetical protein
ORF02383	expressed putative lipoprotein
ORF02393	transporter, putative

E040627 024712

Tabl 10: GBS genes shared with GAS

ORFxxxxx Annotation

ORF02398	transcriptional regulator, Crp/Fnr family
ORF02399	conserved hypothetical protein
ORF02401	acetyltransferase, GNAT family
ORF02403	arginine/ornithine antiporter (arcD)
ORF03002	conserved hypothetical protein, truncation

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF00008 protease, putative
ORF00010 acyl carrier protein (acpP)
ORF00016 acetyltransferase, GNAT family
ORF00018 peptidase, M23/M37 family, putative secreted protein
ORF00035 membrane protein, putative
ORF00087 lipoprotein, putative
ORF00088 hypothetical protein
ORF00089 hypothetical protein
ORF00091 conserved hypothetical protein
ORF00117 ribose ABC transporter, periplasmic D-ribose-binding protein (rbsB)
ORF00118 ribose ABC transporter, permease protein (rbsC)
ORF00120 ribose ABC transporter protein RbsD (rbsD)
ORF00121 ribokinase (rbsK)
ORF00123 hypothetical protein
ORF00130 argininosuccinate lyase (argH)
ORF00137 conserved hypothetical protein
ORF00138 hypothetical protein
ORF00166 4-diphosphocytidyl-2C-methyl-D-erythritol kinase (ispE)
ORF00182 conserved domain protein
ORF00186 transcriptional regulator, Cro/C1 family
ORF00187 hypothetical protein
ORF00188 hypothetical protein
ORF00192 hypothetical protein
ORF00193 conserved hypothetical protein
ORF00196 conserved hypothetical protein
ORF00199 hydrolase, haloacid dehalogenase-like family
ORF00200 sensor histidine kinase, putative
ORF00201 response regulator
ORF00203 conserved hypothetical protein
ORF00204 membrane protein, putative
ORF00205 hypothetical protein
ORF00228 lipoprotein, putative
ORF00234 hypothetical protein
ORF00235 hypothetical protein
ORF00238 hypothetical protein
ORF00240 transcriptional regulator, Cro/C1 family
ORF00241 hypothetical protein
ORF00242 conserved hypothetical protein
ORF00243 hypothetical protein
ORF00244 conserved domain protein
ORF00245 conserved hypothetical protein, fusion
ORF00246 replication initiation protein, putative
ORF00247 hypothetical protein
ORF00248 recombination protein
ORF00249 hypothetical protein
ORF00252 conserved hypothetical protein
ORF00253 hypothetical protein
ORF00254 hypothetical protein
ORF00255 hypothetical protein
ORF00256 hypothetical protein
ORF00257 hypothetical protein
ORF00258 hypothetical protein
ORF00259 hypothetical protein
ORF00260 hypothetical protein
ORF00272 expressed putative lipoprotein

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF00273 hypothetical protein
ORF00274 hypothetical protein
ORF00275 hypothetical protein
ORF00276 hypothetical protein
ORF00278 membrane protein, putative
ORF00279 transcriptional regulator, Cro/Ci family
ORF00280 acetyltransferase, GNAT family
ORF00281 acetyltransferase, GNAT family
ORF00283 conserved hypothetical protein
ORF00284 RNA polymerase sigma factor, ECF subfamily
ORF00285 lipoprotein, putative
ORF00287 transcriptional regulator, TetR family
ORF00288 ABC transporter efflux protein, DrrB family, putative
ORF00292 hypothetical protein
ORF00294 expressed protein of unknown function
ORF00298 acyl carrier protein phosphodiesterase, putative
ORF00308 conserved hypothetical protein
ORF00324 conserved hypothetical protein
ORF00332 hypothetical protein
ORF00340 hypothetical protein
ORF00347 conserved hypothetical protein
ORF00384 hypothetical protein
ORF00402 membrane protein, putative
ORF00408 hypothetical protein
ORF00409 membrane protein, putative
ORF00414 conserved hypothetical protein
ORF00416 hypothetical protein
ORF00417 hypothetical protein
ORF00433 copper-transporter protein CopZ
ORF00448 hypothetical protein
ORF00466 conserved hypothetical protein
ORF00467 acetyltransferase, GNAT family
ORF00475 conserved domain protein
ORF00476 hypothetical protein
ORF00478 carboxymuconolactone decarboxylase family protein
ORF00479 conserved hypothetical protein
ORF00486 transcriptional regulator, AraC family
ORF00487 surface protein Rib
ORF00488 transposase, IS256 family, truncation
ORF00489 DNA-damage-inducible protein J, putative
ORF00490 hypothetical protein
ORF00491 lipoprotein, putative
ORF00493 bacteriophage L54a, integrase, truncation
ORF00497 conserved domain protein
ORF00503 oxidoreductase, Gfo/Idh/MocA family
ORF00506 transposase, IS256 family
ORF00510 bacteriocin transport accessory protein, putative
ORF00512 hypothetical protein
ORF00526 biotin synthetase (bioB)
ORF00527 hypothetical protein
ORF00533 type IV prepilin peptidase-related protein
ORF00538 conserved hypothetical protein
ORF00556 hypothetical protein
ORF00563 expressed protein of unknown function
ORF00575 hypothetical protein

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF00584 conserved hypothetical protein
ORF00585 fructose-1,6-bisphosphatase, putative
ORF00590 carboxymethylenebutenolidase-related protein
ORF00597 conserved hypothetical protein
ORF00598 inosine-uridine preferring nucleoside hydrolase
ORF00599 hypothetical protein
ORF00600 OsmC/Ohr family protein
ORF00608 adenosine deaminase, putative
ORF00610 chorismate mutase, putative
ORF00615 prophage LambdaSa1, site-specific recombinase, phage integrase family
ORF00617 conserved domain protein
ORF00618 hypothetical protein
ORF00620 hypothetical protein
ORF00621 conserved hypothetical protein
ORF00623 hypothetical protein
ORF00624 hypothetical protein
ORF00626 prophage LambdaSa1, transcriptional regulator, Cro/C1 family
ORF00628 hypothetical protein
ORF00630 hypothetical protein
ORF00632 hypothetical protein
ORF00633 conserved hypothetical protein
ORF00635 hypothetical protein
ORF00636 hypothetical protein
ORF00637 hypothetical protein
ORF00638 conserved hypothetical protein
ORF00639 conserved domain protein
ORF00641 prophage LambdaSa1, reverse transcriptase/maturase family protein
ORF00642 conserved hypothetical protein
ORF00643 conserved hypothetical protein
ORF00644 hypothetical protein
ORF00645 hypothetical protein
ORF00646 conserved hypothetical protein
ORF00647 hypothetical protein
ORF00649 hypothetical protein
ORF00650 hypothetical protein
ORF00652 conserved hypothetical protein
ORF00653 conserved hypothetical protein
ORF00657 conserved hypothetical protein, truncation
ORF00661 conserved hypothetical protein
ORF00667 conserved hypothetical protein
ORF00670 prophage LambdaSa1, minor structural protein, putative
ORF00671 prophage LambdaSa1, N-acetylmuramoyl-L-alanine amidase, family 4
ORF00672 prophage LambdaSa1, minor structural protein, putative
ORF00673 hypothetical protein
ORF00674 hypothetical protein
ORF00675 conserved hypothetical protein
ORF00676 conserved hypothetical protein
ORF00678 conserved hypothetical protein
ORF00681 conserved hypothetical protein
ORF00682 hypothetical protein
ORF00683 prophage LambdaSa1, site-specific recombinase, phage integrase family FRAMESHIFT
ORF00685 conserved hypothetical protein
ORF00689 conserved hypothetical protein, FRAMESHIFT
ORF00698 hypothetical protein
ORF00703 phosphoserine phosphatase SerB (serB)

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF00704 MutT/nudix family protein
ORF00712 hypothetical protein
ORF00718 cell wall surface protein, interruption-N
ORF00723 hypothetical protein
ORF00726 transcriptional regulator, AraC family
ORF00727 expressed cell wall surface anchor family protein
ORF00728 expressed cell wall surface anchor family protein
ORF00735 expressed protein of unknown function
ORF00737 conserved hypothetical protein, degenerate
ORF00738 hypothetical protein
ORF00740 hypothetical protein
ORF00741 hypothetical protein
ORF00742 lipoprotein, putative
ORF00747 cylD protein (cylD)
ORF00749 acyl carrier protein AcpC
ORF00750 cylZ protein FRAMESHIFT
ORF00752 cylB protein (cylB)
ORF00753 cylE protein (cylE)
ORF00754 cylF protein (cylF)
ORF00756 cylJ protein (cylJ)
ORF00757 cylK protein (cylK)
ORF00758 hypothetical protein
ORF00759 putative secreted protein
ORF00761 hypothetical protein
ORF00766 expressed putative secreted protein
ORF00767 hypothetical protein
ORF00768 conserved domain protein
ORF00769 permease, putative
ORF00775 conserved hypothetical protein
ORF00777 DedA family protein, putative
ORF00779 membrane protein, putative
ORF00788 sodium:galactoside symporter family protein, putative
ORF00791 transcriptional regulator, GntR family
ORF00793 Glucuronate isomerase (uxaC)
ORF00794 mannonate dehydratase (uxuA)
ORF00795 D-mannonate oxidoreductase
ORF00796 hydrolase, haloacid dehalogenase-like family
ORF00797 glycosyl hydrolase, family 3
ORF00806 conserved hypothetical protein
ORF00822 ABC transporter, ATP-binding protein
ORF00827 hypothetical protein
ORF00834 conserved hypothetical protein
ORF00838 membrane protein, putative
ORF00839 Mn ²⁺ /Fe ²⁺ transporter, NRAMP family
ORF00848 conserved domain protein
ORF00872 cell wall surface anchor family protein
ORF00874 conserved hypothetical protein
ORF00878 ABC transporter, permease protein
ORF00879 YaeC family protein, putative
ORF00888 hydrolase, haloacid dehalogenase-like family
ORF00891 conserved domain protein
ORF00898 conserved hypothetical protein
ORF00900 permease, GntP family
ORF00903 transcriptional regulator, MarR family
ORF00907 glutathion S-transferase family protein

Tabl 11: GBS gene n t shared with GAS or pneumococcus

ORFxxxxx Annotati n
ORF00909 hypothetical protein
ORF00921 membrane protein, putative
ORF00922 glycosyl transferase, family 8
ORF00923 hypothetical protein
ORF00924 conserved hypothetical protein
ORF00939 conserved hypothetical protein
ORF00942 expressed putative secreted protein
ORF00943 hypothetical protein
ORF00944 hypothetical protein
ORF00946 conserved hypothetical protein
ORF00950 hypothetical protein
ORF00951 transcriptional regulator, TenA family
ORF00972 ATP synthase F0, C subunit (atpE)
ORF00980 conserved hypothetical protein
ORF00982 conserved hypothetical protein
ORF01003 conserved hypothetical protein
ORF01004 conserved hypothetical protein
ORF01013 hypothetical protein
ORF01014 hypothetical protein
ORF01015 hypothetical protein
ORF01016 hypothetical protein
ORF01018 hypothetical protein
ORF01019 hypothetical protein
ORF01021 hypothetical protein
ORF01025 HD domain protein
ORF01026 acetyltransferase, GNAT family
ORF01032 chloramphenicol acetyltransferase (cat)
ORF01034 Tn916, transposase
ORF01035 Tn916, excisionase
ORF01037 Tn916, hypothetical protein
ORF01038 Tn916, hypothetical protein
ORF01039 Tn916, transcriptional regulator, putative
ORF01041 Tn916, hypothetical protein
ORF01042 Tn916, NLP/P60 family protein
ORF01044 membrane protein, putative FRAMESHIFT
ORF01048 Tn916, hypothetical protein
ORF01049 Tn916, hypothetical protein
ORF01050 Tn916, hypothetical protein
ORF01051 Tn916, transcriptional regulator, putative
ORF01052 Tn916, FtsK/SpoIIIE family protein
ORF01053 Tn916, hypothetical protein
ORF01054 Tn916, hypothetical protein
ORF01062 hypothetical protein
ORF01086 Na ⁺ /H ⁺ exchanger family protein
ORF01092 acetyltransferase, GNAT family
ORF01096 nisin-resistance protein, putative
ORF01103 conserved hypothetical protein
ORF01124 acetyltransferase, GNAT family
ORF01133 iron-compound ABC transporter, iron-compound-binding protein
ORF01140 conserved hypothetical protein
ORF01142 carbon starvation protein CstA, putative
ORF01143 response regulator
ORF01144 sensor histidine kinase, putative
ORF01145 lipoprotein, putative
ORF01146 conserved hypothetical protein, FRAMESHIFT

Tabl 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF01148 lipoprotein, putative
ORF01149 hypothetical protein
ORF01150 hypothetical protein
ORF01151 hypothetical protein
ORF01152 lipoprotein, putative
ORF01153 hypothetical protein
ORF01157 conserved hypothetical protein
ORF01158 hypothetical protein
ORF01159 hypothetical protein
ORF01160 expressed protein of unknown function FRAMESHIFT
ORF01161 expressed conserved domain protein
ORF01162 conserved hypothetical protein
ORF01164 FtsK/SpoIIIE family protein FRAMESHIFT
ORF01166 hypothetical protein
ORF01167 conserved hypothetical protein
ORF01168 conserved hypothetical protein
ORF01169 hypothetical protein
ORF01172 phage infection protein, putative
ORF01173 conserved hypothetical protein
ORF01174 conserved domain protein
ORF01175 hypothetical protein
ORF01182 membrane protein, putative
ORF01186 cell wall surface anchor family protein, putative
ORF01187 hypothetical protein
ORF01204 hypothetical protein
ORF01215 hypothetical protein
ORF01241 transcriptional regulator, AraC family, putative
ORF01253 rarD protein (rarD)
ORF01257 transporter, BCCT family protein
ORF01258 hypothetical protein
ORF01261 expressed protein of unknown function
ORF01262 conserved hypothetical protein, FRAMESHIFT
ORF01263 hypothetical protein
ORF01265 hypothetical protein
ORF01266 hypothetical protein
ORF01269 conserved hypothetical protein
ORF01272 conserved hypothetical protein
ORF01277 conserved hypothetical protein
ORF01287 conserved hypothetical protein
ORF01288 membrane protein, putative
ORF01299 CMP-N-acetylneuraminic acid synthetase NeuA (neuA)
ORF01300 neuD protein (neuD)
ORF01301 UDP-N-acetylglucosamine-2-epimerase NeuC (neuC)
ORF01302 N-acetyl neuramic acid synthetase NeuB (neuB)
ORF01303 polysaccharide biosynthesis protein CpsL (cpsL)
ORF01304 polysaccharide biosynthesis protein CpsK(V) (cpsK)
ORF01307 glycosyltransferase CpsN(V) (cpsN)
ORF01308 polysaccharide biosynthesis protein CpsM(V) (cpsM)
ORF01309 polysaccharide biosynthesis protein cpsH(V) (cpsH)
ORF01310 glycosyltransferase CpsG(V) (cpsG)
ORF01311 polysaccharide biosynthesis protein CpsF (cpsF)
ORF01312 glycosyltransferase CpsE (cpsE)
ORF01348 conserved domain protein
ORF01349 hypothetical protein
ORF01370 conserved hypothetical protein

Tabl 11: GBS genes not shared with GAS or pn umococcus

ORFxxxxx Ann tati n
ORF01371 conserved hypothetical protein
ORF01372 expressed protein of unknown function
ORF01373 ISSdy1, transposase OrfA
ORF01375 conserved hypothetical protein
ORF01379 transposase OrfB, IS3 family, truncation
ORF01382 GBSi1, group II intron, maturase
ORF01384 hypothetical protein
ORF01385 hypothetical protein
ORF01386 conserved hypothetical protein
ORF01387 conserved hypothetical protein, truncation
ORF01390 ISSdy1, transposase OrfA FRAMESHIFT
ORF01392 hypothetical protein
ORF01393 hypothetical protein
ORF01394 site-specific recombinase, phage integrase family
ORF01395 conserved hypothetical protein
ORF01401 transposase, ISL3 family
ORF01404 mercuric resistance operon regulatory protein MerR (merR)
ORF01408 cadmium efflux system accessory protein (CadC)
ORF01409 conserved hypothetical protein
ORF01410 hypothetical protein
ORF01417 hypothetical protein
ORF01418 hypothetical protein
ORF01420 hypothetical protein
ORF01421 ImpB/MucB/SamB family protein
ORF01423 conserved hypothetical protein
ORF01424 conserved hypothetical protein
ORF01425 conserved hypothetical protein
ORF01426 conserved hypothetical protein
ORF01427 hypothetical protein
ORF01428 conserved hypothetical protein
ORF01430 hypothetical protein
ORF01431 hypothetical protein
ORF01432 conserved domain protein
ORF01433 SNF2 family protein
ORF01434 hypothetical protein
ORF01435 calcium-binding protein, putative
ORF01436 agglutinin receptor (ssp-5)
ORF01437 abortive infection protein AbiGI (abiGI)
ORF01438 abortive infection protein AbiGII (abiGII)
ORF01439 conserved hypothetical protein
ORF01440 expressed protein of unknown function
ORF01441 conserved hypothetical protein, degenerate
ORF01442 membrane protein, putative
ORF01443 hypothetical protein
ORF01444 Tn5252, Orf 21 protein, internal deletion
ORF01445 hypothetical protein
ORF01450 conserved hypothetical protein
ORF01452 hypothetical protein
ORF01454 conserved hypothetical protein
ORF01459 hypothetical protein
ORF01460 homocysteine S-methyltransferase MmuM, putative
ORF01463 hypothetical protein
ORF01464 hypothetical protein
ORF01465 hypothetical protein
ORF01466 transcriptional regulator, TetR family

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF01477 glutathione S-transferase family protein, putative
ORF01478 conserved domain protein
ORF01486 hypothetical protein
ORF01488 R5 protein
ORF01489 transcriptional regulator, MarR family, putative
ORF01494 membrane protein, putative
ORF01497 acetyltransferase, GNAT family
ORF01502 hypothetical protein
ORF01503 conserved hypothetical protein
ORF01508 surface antigen-related protein
ORF01535 conserved hypothetical protein
ORF01547 conserved hypothetical protein
ORF01566 expressed cell wall surface anchor family protein
ORF01572 glycosyltransferase, group 1 family protein
ORF01573 glycosyltransferase, group 2 family protein
ORF01575 membrane protein, putative
ORF01576 glycosyltransferase, group 2 family protein
ORF01577 glycosyltransferase, group 2 family protein
ORF01578 nucleotide sugar dehydratase, putative
ORF01581 lipoprotein, putative
ORF01582 conserved hypothetical protein
ORF01596 ammonium transporter family protein
ORF01597 conserved hypothetical protein
ORF01601 hypothetical protein
ORF01608 proton/peptide symporter family protein
ORF01611 hypothetical protein
ORF01615 conserved domain protein
ORF01638 conserved hypothetical protein
ORF01641 conserved hypothetical protein
ORF01645 cell wall surface anchor family protein
ORF01660 membrane protein, putative
ORF01661 ABC transporter, ATP binding protein
ORF01666 hypothetical protein
ORF01667 hypothetical protein
ORF01670 hypothetical protein
ORF01672 protease, putative, POINT MUTATION
ORF01673 hypothetical protein
ORF01674 hypothetical protein
ORF01675 hypothetical protein
ORF01680 tetracenomycin polyketide synthesis O-methyltransferase TcmP, putative
ORF01681 hypothetical protein
ORF01682 hypothetical protein
ORF01684 hypothetical protein
ORF01692 peptide ABC transporter, ATP-binding protein
ORF01695 peptide ABC transporter, permease protein
ORF01696 peptide ABC transporter, peptide-binding protein
ORF01699 transposase, IS30 family, putative
ORF01700 transporter, major facilitator family
ORF01703 transcriptional regulator, LysR family
ORF01715 conserved hypothetical protein
ORF01719 hypothetical protein
ORF01720 conserved hypothetical protein
ORF01721 glyoxalase family protein
ORF01727 conserved hypothetical protein
ORF01729 acetyltransferase, GNAT family

Tabl 11: GBS g n not shar d with GAS or pn umococcus

ORFxxxxx Annotati n
ORF01730 glycosyl transferase, group 2 family protein
ORF01733 hypothetical protein
ORF01734 conserved hypothetical protein
ORF01735 hypothetical protein
ORF01736 hypothetical protein
ORF01737 hypothetical protein
ORF01742 hypothetical protein
ORF01743 PTS system component, putative
ORF01744 conserved hypothetical protein
ORF01748 D-isomer specific 2-hydroxyacid dehydrogenase family protein
ORF01753 conserved hypothetical protein
ORF01754 hypothetical protein
ORF01761 transposase, IS30 family, putative, truncation
ORF01778 amino acid permease, putative
ORF01807 hypothetical protein
ORF01836 hypothetical protein
ORF01838 hypothetical protein
ORF01839 dihydroxyacetone kinase family protein
ORF01840 transcriptional regulator, TetR family, putative
ORF01842 hypothetical protein
ORF01843 dihydroxyacetone kinase family protein
ORF01844 dihydroxyacetone kinase family protein
ORF01847 conserved hypothetical protein
ORF01850 hypothetical protein
ORF01863 pyruvate phosphate dikinase (ppdK)
ORF01864 expressed protein of unknown function
ORF01865 CBS domain protein
ORF01866 3-hydroxyacyl-CoA dehydrogenase family protein, putative secreted protein
ORF01892 hypothetical protein
ORF01893 hypothetical protein
ORF01894 conserved hypothetical protein
ORF01895 hypothetical protein
ORF01896 hypothetical protein
ORF01897 hypothetical protein
ORF01898 hypothetical protein
ORF01899 hypothetical protein
ORF01903 conserved hypothetical protein
ORF01904 drug resistance transporter, EmrB/QacA family
ORF01905 hypothetical protein
ORF01922 conserved hypothetical protein
ORF01925 FMN-binding protein
ORF01934 hypothetical protein
ORF01936 polyprenyl synthetase family protein
ORF01939 cytochrome d ubiquinol oxidase, subunit II (cydB)
ORF01940 cytochrome d oxidase, subunit I (cydA)
ORF01941 pyridine nucleotide-disulphide oxidoreductase family protein
ORF01942 prenyltransferase, UbiA family
ORF01943 hypothetical protein
ORF01944 hypothetical protein
ORF01946 cyclopropane-fatty-acyl-phospholipid synthase (cfa)
ORF01951 conserved hypothetical protein
ORF01953 hypothetical protein
ORF01954 conserved hypothetical protein
ORF01984 hypothetical protein
ORF01988 hypothetical protein

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF01989 hypothetical protein
ORF01990 hypothetical protein
ORF01991 hypothetical protein
ORF02000 membrane protein, putative
ORF02001 transposase, IS30 family, putative
ORF02005 hypothetical protein
ORF02006 xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (xfp)
ORF02009 conserved hypothetical protein
ORF02010 carbohydrate kinase, FGGY family
ORF02011 hypothetical protein
ORF02012 PTS system component, putative
ORF02015 glyoxylate reductase, NADH-dependent
ORF02016 hypothetical protein
ORF02025 hypothetical protein
ORF02026 hypothetical protein
ORF02030 glutamate-cysteine ligase-related protein
ORF02036 phosphinothricin N-acetyltransferase (pat)
ORF02039 conserved hypothetical protein
ORF02044 conserved hypothetical protein
ORF02045 conserved hypothetical protein
ORF02046 prophage LambdaSa2, lysin, putative
ORF02047 prophage LambdaSa2, holin, putative
ORF02048 conserved hypothetical protein
ORF02049 hypothetical protein
ORF02050 conserved domain protein
ORF02051 prophage LambdaSa2, PblB, putative
ORF02053 conserved hypothetical protein
ORF02056 conserved hypothetical protein
ORF02057 hypothetical protein
ORF02058 hypothetical protein
ORF02059 conserved hypothetical protein
ORF02060 conserved hypothetical protein
ORF02061 hypothetical protein
ORF02062 hypothetical protein
ORF02063 conserved domain protein
ORF02064 conserved domain protein
ORF02066 prophage LambdaSa2, protease, putative
ORF02067 conserved hypothetical protein
ORF02068 prophage LambdaSa2, terminase large subunit, putative
ORF02069 hypothetical protein
ORF02070 hypothetical protein
ORF02071 prophage LambdaSa2, site-specific recombinase, phage integrase family
ORF02072 conserved hypothetical protein
ORF02073 prophage LambdaSa2, transcriptional regulator, Cro/C1 family
ORF02075 hypothetical protein
ORF02077 hypothetical protein
ORF02078 conserved hypothetical protein
ORF02079 conserved hypothetical protein
ORF02080 conserved hypothetical protein
ORF02081 hypothetical protein
ORF02084 prophage LambdaSa2, bacteriophage replication protein/hypothetical protein, truncation/fusion
ORF02085 hypothetical protein
ORF02087 hypothetical protein
ORF02088 conserved hypothetical protein

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF02089 prophage LambdaSa2, HNHendonuclease family protein
ORF02090 prophage LambdaSa2, antirepressor protein, putative
ORF02091 conserved domain protein
ORF02092 hypothetical protein
ORF02093 hypothetical protein
ORF02094 hypothetical protein
ORF02095 prophage LambdaSa2, repressor protein, putative
ORF02097 hypothetical protein
ORF02098 prophage LambdaSa2, site-specific recombinase, phage integrase family
ORF02100 hypothetical protein
ORF02102 hypothetical protein
ORF02103 microcin immunity protein MccF, putative
ORF02105 oxidoreductase, Gfo/Idh/MocA family
ORF02108 hypothetical protein
ORF02109 Cyclic nucleotide-binding domain protein
ORF02119 hypothetical protein
ORF02124 hypothetical protein
ORF02125 nitroreductase family protein
ORF02134 bacteriocin transport accessory protein, putative
ORF02148 neuraminidase-related protein
ORF02160 2',3'-cyclic-nucleotide 2'-phosphodiesterase (cpdB)
ORF02163 conserved hypothetical protein
ORF02171 membrane protein, putative
ORF02172 hypothetical protein
ORF02173 membrane protein, putative
ORF02175 conserved hypothetical protein, truncation
ORF02181 phosphate transport system regulatory protein PhoU, putative
ORF02187 hypothetical protein
ORF02190 conserved hypothetical protein
ORF02191 hypothetical protein
ORF02194 acetyltransferase, GNAT family
ORF02196 hypothetical protein
ORF02198 acetyltransferase, GNAT family
ORF02201 membrane protein, putative
ORF02203 hypothetical protein
ORF02205 transcriptional regulator, Cro/Ci family
ORF02206 conserved hypothetical protein
ORF02207 conserved hypothetical protein TIGR00730
ORF02208 hypothetical protein
ORF02209 site-specific recombinase, phage integrase family
ORF02210 conserved hypothetical protein
ORF02211 conserved hypothetical protein
ORF02212 hypothetical protein
ORF02213 hypothetical protein
ORF02214 transcriptional regulator, Cro/Ci family
ORF02215 expressed protein of unknown function
ORF02216 site-specific recombinase, phage integrase family
ORF02217 conserved hypothetical protein
ORF02219 hypothetical protein
ORF02221 cell wall anchor protein-related protein
ORF02223 hypothetical protein
ORF02224 hypothetical protein
ORF02225 hypothetical protein
ORF02226 membrane protein, putative
ORF02227 conjugal transfer protein, interruption-C

Tabl 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF02230 conserved hypothetical protein
ORF02231 conserved hypothetical protein
ORF02232 conserved hypothetical protein
ORF02235 hypothetical protein
ORF02236 conserved hypothetical protein
ORF02237 hypothetical protein
ORF02238 hypothetical protein
ORF02239 hypothetical protein
ORF02240 transcriptional regulator, Cro/Ci family
ORF02241 hypothetical protein
ORF02242 transcriptional regulator, Cro/Ci family
ORF02243 FtsK/SpoIIIE family protein
ORF02244 hypothetical protein
ORF02245 hypothetical protein
ORF02246 cell wall surface anchor family protein
ORF02247 transposase, ISL3 family
ORF02250 mercuric resistance operon regulatory protein MerR (merR)
ORF02251 Mn2+/Fe2+ transporter, NRAMP family
ORF02252 membrane protein, putative
ORF02253 ABC transporter, ATP-binding protein
ORF02254 conserved hypothetical protein
ORF02255 streptomycin resistance protein
ORF02257 hypothetical protein
ORF02258 hypothetical protein
ORF02259 conserved hypothetical protein
ORF02260 acetyltransferase, GNAT family
ORF02261 membrane protein, putative
ORF02263 hypothetical protein
ORF02264 transcriptional regulator, Cro/Ci family
ORF02265 PAP2 family protein
ORF02266 conserved hypothetical protein FRAMESHIFT
ORF02267 conserved hypothetical protein TIGR00730
ORF02268 protease, putative
ORF02269 rhodanese family protein
ORF02271 hypothetical protein
ORF02274 conserved hypothetical protein
ORF02275 5-methyltetrahydrofolate--homocysteine methyltransferase, putative
ORF02277 conserved hypothetical protein
ORF02279 hypothetical protein
ORF02282 sensor histidine kinase
ORF02283 chromosome assembly-related protein
ORF02287 expressed protein of unknown function
ORF02291 pathogenicity protein, putative
ORF02308 hydrolase, haloacid dehalogenase-like family
ORF02314 conserved hypothetical protein
ORF02317 hypothetical protein
ORF02330 hypothetical protein
ORF02344 site-specific recombinase, phage integrase family
ORF02345 conserved hypothetical protein
ORF02346 conserved hypothetical protein
ORF02347 hypothetical protein
ORF02349 conserved hypothetical protein
ORF02350 hypothetical protein
ORF02351 transcriptional regulator, Cro/Ci family
ORF02352 conserved domain protein

Table 11: GBS genes not shared with GAS or pneumococcus

ORFxxxxx Annotation
ORF02354 hypothetical protein
ORF02356 expressed putative secreted protein
ORF02362 sensor histidine kinase
ORF02363 response regulator
ORF02367 membrane protein, putative
ORF02368 conserved hypothetical protein
ORF02379 membrane protein, putative
ORF02395 transcriptional regulator, Cro/Ci family
ORF02406 membrane protein, putative
ORF02416 diacylglycerol kinase catalytic domain protein, putative
ORF02418 hypothetical protein
ORF02422 hypothetical protein
ORF02425 conserved hypothetical protein
ORF03001 conserved hypothetical protein
ORF03004 conserved hypothetical protein
ORF03005 cylX protein
ORF03006 Tn916, hypothetical protein
ORF03007 Tn916, hypothetical protein
ORF03008 Tn916, hypothetical protein
ORF03009 Tn916, tetM leader peptide
ORF03010 Tn916, hypothetical protein
ORF03012 prophage LambdaSa2, HNH endonuclease family protein
ORF03013 conserved hypothetical protein
ORF03015 conjugal transfer protein, Interruption-N

Table 12: GBS ORF's not shared with GAS, pneumococcus or any published genome

ORFxxxxx Annotation

ORF00035	membrane protein, putative
ORF00087	lipoprotein, putative
ORF00088	hypothetical protein
ORF00089	hypothetical protein
ORF00123	hypothetical protein
ORF00138	hypothetical protein
ORF00187	hypothetical protein
ORF00188	hypothetical protein
ORF00192	hypothetical protein
ORF00205	hypothetical protein
ORF00228	lipoprotein, putative
ORF00234	hypothetical protein
ORF00235	hypothetical protein
ORF00238	hypothetical protein
ORF00240	transcriptional regulator, Cro/C1 family
ORF00241	hypothetical protein
ORF00242	conserved hypothetical protein
ORF00243	hypothetical protein
ORF00247	hypothetical protein
ORF00249	hypothetical protein
ORF00253	hypothetical protein
ORF00254	hypothetical protein
ORF00255	hypothetical protein
ORF00256	hypothetical protein
ORF00257	hypothetical protein
ORF00258	hypothetical protein
ORF00259	hypothetical protein
ORF00260	hypothetical protein
ORF00272	expressed putative lipoprotein
ORF00273	hypothetical protein
ORF00274	hypothetical protein
ORF00275	hypothetical protein
ORF00276	hypothetical protein
ORF00278	membrane protein, putative
ORF00285	lipoprotein, putative
ORF00292	hypothetical protein
ORF00294	expressed protein of unknown function
ORF00308	conserved hypothetical protein
ORF00332	hypothetical protein
ORF00340	hypothetical protein
ORF00384	hypothetical protein
ORF00402	membrane protein, putative
ORF00408	hypothetical protein
ORF00416	hypothetical protein
ORF00417	hypothetical protein
ORF00448	hypothetical protein
ORF00476	hypothetical protein
ORF00489	DNA-damage-inducible protein J, putative
ORF00490	hypothetical protein
ORF00491	lipoprotein, putative
ORF00497	conserved domain protein
ORF00510	bacteriocin transport accessory protein, putative
ORF00512	hypothetical protein
ORF00527	hypothetical protein
ORF00556	hypothetical protein

Tabl 12: GBS ORF's not shared with GAS, pneumococcus r any published genom

ORFxxxxx Ann tati n
ORF00575 hypothetical protein
ORF00599 hypothetical protein
ORF00618 hypothetical protein
ORF00620 hypothetical protein
ORF00623 hypothetical protein
ORF00626 prophage LambdaSa1, transcriptional regulator, Cro/Ci family
ORF00628 hypothetical protein
ORF00630 hypothetical protein
ORF00632 hypothetical protein
ORF00635 hypothetical protein
ORF00636 hypothetical protein
ORF00637 hypothetical protein
ORF00642 conserved hypothetical protein
ORF00644 hypothetical protein
ORF00645 hypothetical protein
ORF00647 hypothetical protein
ORF00649 hypothetical protein
ORF00650 hypothetical protein
ORF00653 conserved hypothetical protein
ORF00657 conserved hypothetical protein, truncation
ORF00661 conserved hypothetical protein
ORF00673 hypothetical protein
ORF00674 hypothetical protein
ORF00675 conserved hypothetical protein
ORF00676 conserved hypothetical protein
ORF00682 hypothetical protein
ORF00685 conserved hypothetical protein
ORF00698 hypothetical protein
ORF00712 hypothetical protein
ORF00718 cell wall surface protein, interruption-N
ORF00723 hypothetical protein
ORF00735 expressed protein of unknown function
ORF00737 conserved hypothetical protein, degenerate
ORF00738 hypothetical protein
ORF00740 hypothetical protein
ORF00741 hypothetical protein
ORF00747 cylD protein (cylD)
ORF00753 cylE protein (cylE)
ORF00756 cylJ protein (cylJ)
ORF00757 cylK protein (cylK)
ORF00758 hypothetical protein
ORF00759 putative secreted protein
ORF00761 hypothetical protein
ORF00796 hydrolase, haloacid dehalogenase-like family
ORF00806 conserved hypothetical protein
ORF00822 ABC transporter, ATP-binding protein
ORF00827 hypothetical protein
ORF00872 cell wall surface anchor family protein
ORF00909 hypothetical protein
ORF00923 hypothetical protein
ORF00924 conserved hypothetical protein
ORF00942 expressed putative secreted protein
ORF00943 hypothetical protein
ORF00944 hypothetical protein
ORF01013 hypothetical protein

Table 12: GBS ORF's not shared with GAS, pneumococcus or any published genome

ORFxxxxx Annotation

ORF01014 hypothetical protein
ORF01015 hypothetical protein
ORF01016 hypothetical protein
ORF01018 hypothetical protein
ORF01019 hypothetical protein
ORF01021 hypothetical protein
ORF01035 Tn916, excisionase
ORF01062 hypothetical protein
ORF01096 nisin-resistance protein, putative
ORF01145 lipoprotein, putative
ORF01146 conserved hypothetical protein, FRAMESHIFT
ORF01148 lipoprotein, putative
ORF01149 hypothetical protein
ORF01150 hypothetical protein
ORF01151 hypothetical protein
ORF01152 lipoprotein, putative
ORF01153 hypothetical protein
ORF01158 hypothetical protein
ORF01159 hypothetical protein
ORF01161 expressed conserved domain protein
ORF01162 conserved hypothetical protein
ORF01166 hypothetical protein
ORF01168 conserved hypothetical protein
ORF01169 hypothetical protein
ORF01174 conserved domain protein
ORF01175 hypothetical protein
ORF01186 cell wall surface anchor family protein, putative
ORF01187 hypothetical protein
ORF01204 hypothetical protein
ORF01215 hypothetical protein
ORF01258 hypothetical protein
ORF01262 conserved hypothetical protein, FRAMESHIFT
ORF01263 hypothetical protein
ORF01265 hypothetical protein
ORF01266 hypothetical protein
ORF01304 polysaccharide biosynthesis protein CpsK(V) (cpsK)
ORF01308 polysaccharide biosynthesis protein CpsM(V) (cpsM)
ORF01309 polysaccharide biosynthesis protein cpsH(V) (cpsH)
ORF01349 hypothetical protein
ORF01384 hypothetical protein
ORF01385 hypothetical protein
ORF01386 conserved hypothetical protein
ORF01392 hypothetical protein
ORF01395 conserved hypothetical protein
ORF01409 conserved hypothetical protein
ORF01410 hypothetical protein
ORF01417 hypothetical protein
ORF01418 hypothetical protein
ORF01420 hypothetical protein
ORF01423 conserved hypothetical protein
ORF01424 conserved hypothetical protein
ORF01425 conserved hypothetical protein
ORF01426 conserved hypothetical protein
ORF01427 hypothetical protein
ORF01431 hypothetical protein

Tabl 12: GBS ORF's not shared with GAS, pneum coccus r any published genome

ORFxxxxx Annotation

ORF01432 conserved domain protein
ORF01434 hypothetical protein
ORF01435 calcium-binding protein, putative
ORF01437 abortive infection protein AbiGI (abiGI)
ORF01438 abortive infection protein AbiGII (abiGII)
ORF01441 conserved hypothetical protein, degenerate
ORF01443 hypothetical protein
ORF01445 hypothetical protein
ORF01452 hypothetical protein
ORF01459 hypothetical protein
ORF01463 hypothetical protein
ORF01464 hypothetical protein
ORF01465 hypothetical protein
ORF01486 hypothetical protein
ORF01488 R5 protein
ORF01575 membrane protein, putative
ORF01581 lipoprotein, putative
ORF01601 hypothetical protein
ORF01611 hypothetical protein
ORF01638 conserved hypothetical protein
ORF01645 cell wall surface anchor family protein
ORF01660 membrane protein, putative
ORF01666 hypothetical protein
ORF01667 hypothetical protein
ORF01670 hypothetical protein
ORF01673 hypothetical protein
ORF01674 hypothetical protein
ORF01675 hypothetical protein
ORF01681 hypothetical protein
ORF01682 hypothetical protein
ORF01684 hypothetical protein
ORF01719 hypothetical protein
ORF01733 hypothetical protein
ORF01735 hypothetical protein
ORF01736 hypothetical protein
ORF01737 hypothetical protein
ORF01742 hypothetical protein
ORF01754 hypothetical protein
ORF01761 transposase, IS30 family, putative, truncation
ORF01807 hypothetical protein
ORF01836 hypothetical protein
ORF01838 hypothetical protein
ORF01842 hypothetical protein
ORF01850 hypothetical protein
ORF01892 hypothetical protein
ORF01893 hypothetical protein
ORF01895 hypothetical protein
ORF01896 hypothetical protein
ORF01897 hypothetical protein
ORF01898 hypothetical protein
ORF01899 hypothetical protein
ORF01905 hypothetical protein
ORF01934 hypothetical protein
ORF01943 hypothetical protein
ORF01944 hypothetical protein

Table 12: GBS ORF's not shared with GAS, pneumococcus or any published genome

ORFxxxxx Annotation

ORF01953 hypothetical protein
ORF01984 hypothetical protein
ORF01988 hypothetical protein
ORF01989 hypothetical protein
ORF02005 hypothetical protein
ORF02011 hypothetical protein
ORF02016 hypothetical protein
ORF02025 hypothetical protein
ORF02026 hypothetical protein
ORF02045 conserved hypothetical protein
ORF02047 prophage LambdaSa2, holin, putative
ORF02048 conserved hypothetical protein
ORF02049 hypothetical protein
ORF02050 conserved domain protein
ORF02053 conserved hypothetical protein
ORF02057 hypothetical protein
ORF02058 hypothetical protein
ORF02061 hypothetical protein
ORF02062 hypothetical protein
ORF02063 conserved domain protein
ORF02067 conserved hypothetical protein
ORF02069 hypothetical protein
ORF02070 hypothetical protein
ORF02072 conserved hypothetical protein
ORF02073 prophage LambdaSa2, transcriptional regulator, Cro/C1 family
ORF02075 hypothetical protein
ORF02077 hypothetical protein
ORF02078 conserved hypothetical protein
ORF02081 hypothetical protein
ORF02085 hypothetical protein
ORF02087 hypothetical protein
ORF02088 conserved hypothetical protein
ORF02091 conserved domain protein
ORF02092 hypothetical protein
ORF02093 hypothetical protein
ORF02094 hypothetical protein
ORF02097 hypothetical protein
ORF02100 hypothetical protein
ORF02102 hypothetical protein
ORF02108 hypothetical protein
ORF02119 hypothetical protein
ORF02124 hypothetical protein
ORF02171 membrane protein, putative
ORF02172 hypothetical protein
ORF02173 membrane protein, putative
ORF02191 hypothetical protein
ORF02196 hypothetical protein
ORF02203 hypothetical protein
ORF02208 hypothetical protein
ORF02212 hypothetical protein
ORF02213 hypothetical protein
ORF02214 transcriptional regulator, Cro/C1 family
ORF02215 expressed protein of unknown function
ORF02217 conserved hypothetical protein
ORF02219 hypothetical protein

**Tabl 12: GBS ORF' not shared with GAS, pn um coccus or any publish d
g n m**

ORFxxxxx Ann tati n

ORF02221 cell wall anchor protein-related protein
ORF02223 hypothetical protein
ORF02224 hypothetical protein
ORF02225 hypothetical protein
ORF02231 conserved hypothetical protein
ORF02235 hypothetical protein
ORF02236 conserved hypothetical protein
ORF02237 hypothetical protein
ORF02238 hypothetical protein
ORF02239 hypothetical protein
ORF02241 hypothetical protein
ORF02244 hypothetical protein
ORF02245 hypothetical protein
ORF02263 hypothetical protein
ORF02268 protease, putative
ORF02271 hypothetical protein
ORF02279 hypothetical protein
ORF02283 chromosome assembly-related protein
ORF02317 hypothetical protein
ORF02330 hypothetical protein
ORF02344 site-specific recombinase, phage integrase family
ORF02345 conserved hypothetical protein
ORF02347 hypothetical protein
ORF02349 conserved hypothetical protein
ORF02350 hypothetical protein
ORF02351 transcriptional regulator, Cro/C1 family
ORF02354 hypothetical protein
ORF02356 expressed putative secreted protein
ORF02395 transcriptional regulator, Cro/C1 family
ORF02418 hypothetical protein
ORF02422 hypothetical protein
ORF02425 conserved hypothetical protein
ORF03004 conserved hypothetical protein
ORF03005 cyIX protein
ORF03006 Tn916, hypothetical protein
ORF03007 Tn916, hypothetical protein
ORF03008 Tn916, hypothetical protein
ORF03009 Tn916, tetM leader peptide
ORF03010 Tn916, hypothetical protein
ORF03015 conjugal transfer protein, interruption-N

Table 13: Comparative Sequences relating to SAG0466 (thiolase)

SEQ ID NO. 1301: SAG0466 FROM THE 2603V/R GBS STRAIN

CTCCTGCCCTGCAATGGCAGTTAGACCCATAGGTTTATTTTATATTTTAAATGCCTGCATAAGATGAAGGATATTAATA
 ATTCTGAGCAGGCATAAGGGTGTCCGTAAGCTAATGTCCCTCCAAAAATATTGAATTTTCTCTCTCTCAGGATAATA
 ATGATTAAATAGAGCATCAATCGCTGCAATGGTTCATTCCATTCAATTGCATCATAATCCGATATTTTAGTATGAGTTT
 CTGTTAATAGTTTTTCCGTAGCCGTGTGAACCAATTCGGACTAAGCTTGGGATCTCCTGCTACTTCTACAATGTGAACA
 ATCCGGAATTCGTTTTCTGACTCTGAAGCGTTAGAAATGCAGCAGCATCGTGCATTAAACAAACATTTCCAATAGTGAG
 CAAAGGTGAATTTTCCATCAATCTTGGTAATTTTTGAAAAATGTTtCTTTaGTTTTCTAACGCCCTTGATCTCGCATCC
 CTTCCATTGGTAAGATTACyTCTTCTAAATAGCCACCTTGTTTAGCTGTTAAGGCGCGTTTATGGCTCAAGAATGCCAAT
 TTATCTAACATTTCTCTTCTAAaCCATATTTTTGACAGACTCTCTGGGCCCCCTTCTAACATTACAGTTTCAGCATAAGA
 GTCAGGAGAAAACCTGAGCAACTGTATATTCTCCGTTACGATTATCTTCTTTAGCATAACGCTCTCATAGGTTGAAGAGAAC
 TACTTTCAATCCCCCAACAAGAATTTTTTCATTAAATACCGGTACTGATTTTTAGATAACCAAAAAACAAGGCAGAACTT
 GATGAAGCACACTGCATATCAATCGTTTGTACTGGAATATAGGATTATAATCAGAAAAAGAGTCATCAAACGACCAAT
 ATTGCCCCAGTACCAACTGTGTTCCACAAATAATACTATCAATGTTAGATTCTGATTCTATTTTTTTTTATTGATTTA
 AAAGGTGTGCTCCTAAAGTTCTGGACGGTAAGTTTAAATTGCTT

SEQ ID NO. 1302: SAG0466 FROM THE M732 GBS TYPE III STRAIN

TCGGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAATAAAAAAATA
 GAATCAGAATCTAATATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCT
 TTTTCTGATTATGAATCCTATATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGTTCTGCCTTGTTTTTG
 GTTATCTAAAAATCAGTGCCGGTATTAATGAAAAAGTCTTGTGGGGGATTGAAAGTAGTTCTCTTCAACCTATGAGA
 CGTTACGCTAAAGAAGATAATCGTAACGGAGAAATATACCGTTGCTCAGTTTTCTCCTGACTCTTATGCTGAACTGTAAT
 GTTAGAAGGGGACAAAGAGTCTGTCAAAAATATGGTTTAGAAGAGAAATGTTAGATAAATTGGCATTCTTGAGCCATA
 AACGCGCCTTAACAGCTAAACAAGGTGCTATTTAGAAGAGGTAATCTTACCAATGGAAGGGATGCGAGATCAAGGCGTT
 AGAAAACATAAAGAAGCATTTTTTCAAAAATTACCAAGATTGATGGAAAATTCACCTTTGCTCACTATTGGAAATGTTG
 TTTAATGCACGATGCTGCTGCATTCTAACGCTTCAGAGTCAGAAAACAGAATTCGGGATTGTTACATTGTAGAAGTAG
 CAGGAGATCCCAAGCTTAGTCCAGAATTGGTTACACGGCTACGGAAAACTATTAACAGAACTCATACTAAAATATCG
 GATTATGATGCAATTGAATGGAATGAACCATTTCAGCGATTGATGCTTTATTTAATCATTTATCTCTGAAGAGAGAGA
 AAAATTCAATATTTTTGGAGGGGCATTAGCTTACGGACACCCTTATGCCTGCTCAGGAATTA

SEQ ID NO. 1303: SAG0466 FROM THE 090 GBS TYPE Ia STRAIN

TTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCTTTTTCTGATTATGAATCCTATATTCCAG
 TACAAACGATTGATATGCAGTGTGCTTCATCAAGTTCTGCCTTGTTTTTGGTTATCTAAAAATCAGTGCCGGTATTAAT
 GAAAAAGTTCTTGTGGGGGGATTGAAAGTAGTTCTCTTCAACCTATGAGACGTTACGCTAAAGAAGATAATCGTAACGG
 AGAATATACCGTTGCTCAGTTTTCTCCTGACTCTTAKGCTGAACTGTAATGtTAGAAGGGGACAAAGAGTCTGTCAA
 AATATGGTTTTAGAAGAGAAATGTTAGATAAATTGGCATTCTTTAGGCCATAAACCGCGCCTTAACAGCTAAACAAGGTGGC
 TATTTAGAAGAGGTAATCTTACCAATGGAAGGGATGCGAGATCAAGGCGTTAGAAAACATAAAGAAGCATTTTTTCAAAA
 ATTACCAAGATTGATGGrAAATTCACCTTTGCTCACTATTGGAAATGTTTGTTAATGCACGATGCTGCTGCATTTCTwa
 CGCTTCAGAGTCAGAAAACAGAATTCGGGATTGTTACATTGTAGAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTG
 GTTCACACGGCTACGGAAAACTATTAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACC
 ATTTGCAGCGATTGATGCTTTATTTAATCATTATTATCTCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGGCATTAG
 CTTACGGACACCCTTATGCCTGCTCAGG

SEQ ID NO. 1304: SAG0466 FROM THE COH1 GBS TYPE Ia STRAIN

ATCGGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAATAAAAAAATA
 GAATCAGAATCTAATATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCT
 TTTTCTGATTATGAATCCTATATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGTTCTGCCTTGTTTTTG
 GGTATCTAAAA

SEQ ID NO. 1305 : SAG0466 FROM THE CJB GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

TTTTCAAAAATTACCAAGATTGATGGAAAATTCACCTTTGCTCACTATTGGAAATGTTTGTTAATGCACGATGCTGCTG
 CATTTCTAACGCTTCAGAGTCAGAAAACAGAATTCGGGATTGTTACATTGTAGAAGTAGCAGGAGATCCCAAGCTTAGT
 CCAGAATTGGTTACACGGCTACGGAAAACTATTAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATG
 GAATGAACCATTTCAGCGATTGATGCTTTATTTAATCATTATTATCTCTGAAGAGAGAGAAAAATTCAATATTTTTGGAG
 GGGCATTAGCTTACGGACACCCTTAATGCCTGCTCAGGAATTATTAATATCC

SEQ ID NO. 1306: sag0466 FROM THE CJB110 GBS NONTYPEABLE STRAIN

GGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAATAAAAAAATA
 ACCAGAATCTAACATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCTT
 TTTCTGATTATGAATCCTATATTC

Table 13: Comparative Sequences relating to SAG0466 (thiolase)

SEQ ID NO. 1307: SAG0466 FROM THE 1169NT1 GBS TYPE V STRAIN REVERSE COMPLEMENT
CAAGATTGATGGAAAATTCACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTCTAACGCTT
CAGAGTCAGAAAACAGAATTCGGGATTGTTACATTGTAGAAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTCA
CACGGCTACGGAAAACTATTAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTG
CAGCGATTGATGCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCATATTTTGGAGGGGCATTAGCTTAC
GGACACCTTATGCCTGCTCAGGAATTATTAATATCCTTCATCTTATGCAGGCATTAAAAATATAAAAATAAACCTATGGG
CCTAACTGCCATTGCAGGGGCA

SEQ ID NO. 1308: SAG0466 FROM THE 18RS21 GBS TYPE II STRAIN
CCTTAACAGTTAAACAAGGTGGCTATTTAGAAGAGGTAATCTTACCAATGGAAGGGATGCGAGATCAAGGCGTTAGAAAA
CTAAAGAAACATTTTTTCAAAAATTACCAAGATTGATGGAAAATTCACCTTTGCTCACTATTGGAAATGTTTGTGTTAAT
GCACGATGCTGCTGCATTCTAACGCTTCAGAGTCAGAAAACAGAATTCGGGATTGTTACATTGTAGAAAGTAGCAGGAG
ATCCCAAGCTTAGTCCAGAATTGGTTACACGGCTACGGAAAACTATTAACAGAACTCATACTAAAATATCGGATTAT
GATGCAATTGAATGGAATGAACCATTTGCAGCGATTGATGCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAAT
CAATATTTTTGGAGGGACATTAGCTTACGGACACCTTATGCCTGCTCAGGAATTATTAATATCCTTCATCTTATGCAGG
CATTAAAAATATAAAAATAAACCTATGGGTCTAACTGCCATTGCAGGGGCA

SEQ ID NO. 1309: SAG0466 FROM THE 18RS21 GBS TYPE II STRAIN
TCGGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTTAAATCAAATAAAAAAATA
GAATCAGAATCTAACATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCT
TTTTTCTGATTATGAATCCTATATTTCCAGTACAAACGATTGATATGCTGAGTGTGCTTCATCAAGTTCTGCCTTGTGTTT
GTTATCTAAAAATCAGTACCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAAAGTAGTTCTCTCAACCTATGAGA
CGTTATGCTAAAGAAGATAATCGTAACGGAGAATATACAGTTGCTCAGTTTTCTCCTGACTCTTATGCTGAAACTGTAAT
GTTAGAAGGGGCCAGAGAGTCTGTCAAAAATATGGTTTTAGAAGAGAAATGTTAGATAAATTGGCATTCTTGAGCCATA
AACCGCCTTAACAGCTAAACA

SEQ ID NO. 1310: SAG0466 FROM THE H36b GBS TYPE Ib STRAIN
TTTGGGCTACGAACACCTATCGGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTT
AAATCAAATAAAAAAATAGAATCAGAATCTAACATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATA
TTGGTCGTTTGATGACTCTTTTTCTGATTATGAATCCTATATTTCCAGTACAAACGATTGATATGCAAGTGTGCTTCATCA
AGTTCTGCCTTGTGTTTTTGGTTATCTAAAAATCAGTACCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAAAGTAG
TTCTCTTCAACCTATGAGACGTTATGCTAAAGAAGATAATCGTAACGGAGAATATACAGTTGCTCAGTTTTCTCCTGACT
CTTATGCTGAAACTGTAATGTTAGAAGGGGCC

SEQ ID NO. 1311: SAG0466 FROM THE H36b GBS TYPE Ib STRAIN (REVERSE COMPLEMENT)
GAAAATTCACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTCTAACGCTTCAGAGTCAGAA
AACAGAATTCGGGATTGTTACATTGTAGAAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTACACGGCTACGG
AAAACTATTAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGAT
GCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCATATTTTGGAGGGACATTAGCTTACGGACACCTTA
TGCCTGCTCAGGAATTATTAATATCCTTCATCTTATGCAGGCATTAAAAATATAAAAATAAACCTATGGGTCTAACTGCCA
TTGCAGGGGCAGGA

SEQ ID NO. 1312: SAG0466 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
CCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTCTAACGCTTCAGAGTCAGAAAACAGAATT
CCGGATTGTTACATTGTAGAAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTACACGGCTACGGAAAACTAT
TAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGATGCTTTATTT
AATCATTATTATCCTGAAGAGAGAGAAAAATTCATATTTTGGAGGGGCATTAGCTTACGGACACCTTATGCCTGCTC
AGGAATTATTAATATCCTTCATCTTATGCAGGCATTAAAAATATAAAAATAAACCTATGGGTCTAACTGCG

SEQ ID NO. 1313: SAG0466 FROM THE M781 GBS TYPE III STRAIN
GCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAATAAAAAAATAGAATCAGAATCTAATA
TTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATATTGGTCGTTTGATGACTCTTTTTTCTGATTATGAA
TCCTATATTTCCAGTACAAACGATTGATATGCAAGTGTGCTTCATCAAGTTCTGCCTTGTGTTTTTGGTTATCTAAAAATCAG
TGCCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAAAGTAGTTCTCTTCAACCTATGAGACGTTACGGCTAAAGAAG
ATAATCGTAACGGAGAATATACCGTTGCTCAGTTTTCTCCTGACTCTTATGCTGAACTGTAATGTTAGA

SEQ ID NO 1314: SAG0466 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)
CCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTCTAACGCTTCAGAGTCAGAAAACAGAATT
CCGGATTGTTACATTGTAGAAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTACACGGCTACGGAAAACTAT
TAACAGAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGATGCTCTATTT
AATCATTATTATCCTGAAGAGAGAGAAAAATTCATATTTTGGAGGGACATTAGCTTACGGACACCTTATGCCTGCTC
AGGAATTATTAATATCCTTCATCTTATGCAGGCATTAAAAATATAAAAATAAACCTATGGGTCTAACTGCCATTGCAGGGG
C

Table 13: Comparative Sequences relating to SAG0466 (phage)

SEQ ID NO. 1315: SAG0466 FROM THE JM9130013 GBS TYPE VIII STRAIN REVERSE
COMPLEMENT

GCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTCTAACGCTTCAGAGTCAGAAAAACAGAATTCGGGA
TTGTTTCACATTGTAGAAGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTTCACACGGCTACGGAAAAACTATTAAACA
GAAACTCATACTAAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTCAGCGATTGATGCTCTATTTAATCA
TTATTATCCTGAAGAGAGAGAAAAATTCATATTTTGGAGGGGCATTAGCTTACGGACACCCTTATGCCTGCTCAGGAA
TTATTAATATCCTTCATCTTATGCAGGCATTAAAAATATAAAATAAACCTATGGGTCTAACTGCCATTGCAGGGGCAGGA

SEQ ID NO. 1316: SAG0466 FROM THE JM9130013 GBS TYPE VIII STRAIN

TTTGGGCTACGAACACCTATCGGTATAAAAGGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTT
AAATCAATAAAAAAATAGAAATCAGAAATCTAACATTGATAGTATTATTTGTGGGAACACAGTTGGTACTGGGGGCAATA
TTGGTCGTTTGTGACTCTTTTTCTGATTATGAATCCTATATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCA
AGTTCTGCCTTGTTTTTTGGTTATCTAAAAATCAGTACCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAAAGTAG
TTCTCTTCAACCTATGAGACGTTATGCTAAAGAAGATAATCGTAACGGAGAATATA

SEQ1301 -----CTCCTGCCCCCTGCAATGGCAGTTAGACCCATAGGTTTATTTTTATATTTTA
SEQ1302 -----
SEQ1303 -----
SEQ1304 -----
SEQ1305 -----
SEQ1306 -----
SEQ1307 -----
SEQ1308 CTTAACAGTTAAACAAGGTGGCTATTTAGAAGAGGTAATCTTACCAATGGAAGGGATGC
SEQ1309 -----
SEQ1310 -----
SEQ1311 -----
SEQ1312 -----
SEQ1313 -----
SEQ1314 -----
SEQ1315 -----
SEQ1316 -----

SEQ1301 TGCCTGCATAAGATGAAGGATATTAATAATTCTGAGCAGGCATAAGGGTGTCCGTAAG
SEQ1302 -----TCGGTATAAA
SEQ1303 -----
SEQ1304 -----ATCGGTATAAA
SEQ1305 -----TTTTCAAAAATTACCAAGATTGATGG
SEQ1306 -----GGTATAAA
SEQ1307 -----CAAGATTGATGG
SEQ1308 AGATCAAGGCGTTAGAAAACTAAAAAGAAACATTTTTTTCAAAAATTACCAAGATTGATGG
SEQ1309 -----TCGGTATAAA
SEQ1310 -----TTTGGGCTACGAACACCTATCGGTATAAA
SEQ1311 -----G
SEQ1312 -----
SEQ1313 -----
SEQ1314 -----
SEQ1315 -----
SEQ1316 -----TTTGGGCTACGAACACCTATCGGTATAAA

SEQ1301 TAATGTCCCTCCAAA-AATATTGAATTTTCTCTCTC-TTCAGGATAATAATGATTAA
SEQ1302 GGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAAT
SEQ1303 -----
SEQ1304 GGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAAT
SEQ1305 AAATTACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1306 GGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAAT
SEQ1307 AAATTACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1308 AAATTACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1309 GGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTTAAATCAAAT
SEQ1310 GGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTTAAATCAAAT
SEQ1311 AAATTACCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1312 -----CCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1313 -----GCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTCTTAAATCAAAT
SEQ1314 -----CCTTTGCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT
SEQ1315 -----GCTCACTATTGGAAATGTTTGTGTTAATGCACGATGCTGCTGCATTTT

Table 13: Comparative Sequences relating to SAG0466 (Polase)

SEQ1316	GGGAAGCAATTTAAACATTACCGTCCAGAACTTTTAGGAGCACACCTTTTAAATCAAAT
SEQ1301	AGAGCATCAATCGCTGCAATGGTTCATTCC-ATTCAATTGCATCATAATCCGATATTT
SEQ1302	AAAAAATAGAAATCAGAATCTAAT--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1303	-----TTGTGGGAACA-CAGT
SEQ1304	AAAAAATAGAAATCAGAATCTAAT--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1305	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1306	AAAAAATATAACCAGAATCTAAC--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1307	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1308	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1309	AAAAAATAGAAATCAGAATCTAAC--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1310	AAAAAATAGAAATCAGAATCTAAC--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1311	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1312	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1313	AAAAAATAGAAATCAGAATCTAAT--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1314	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1315	AACGCTTCAGAGTCAGAAAACAGA--ATTCCGGATTGTTTCACATTGTAGAAGTAGCAGG
SEQ1316	AAAAAATAGAAATCAGAATCTAAC--ATT---GATAGTATTATTTGTGGGAACA-CAGT
SEQ1301	AGTATGAGTTTCTGTTAATAGTTTTTCCGTAGCCGTGTGAACCAATTCTGGACTAAGCT
SEQ1302	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1303	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1304	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1305	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1306	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1307	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1308	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1309	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1310	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1311	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1312	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1313	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1314	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1315	GATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAAACTATTAACAGAAAC
SEQ1316	GGTACTGGGGGCAATATTGG-TCGTTTGATGACTCTTTTTTCTGATTATGAATCCTA--
SEQ1301	GGGATCTCCTGCTACTTCTACAATGTGAACAATCCGGA-ATTCTGTTTTCTGACTCTGA
SEQ1302	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1303	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1304	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1305	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1306	TATTC-----
SEQ1307	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1308	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1309	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1310	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1311	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1312	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1313	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT
SEQ1314	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1315	CATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTTGCAGCGATTGA
SEQ1316	TATTCCAGTACAAACGATTGATATGCAGTGTGCTTCATCAAGT--TCTGCCTTGTTTTT

444444 2 002704

Table 13: Comparative Sequences relating to SAG0466 (base)

SEQ1301	GC GTTAGAAATGCAGCAGCATCGTGCATTAAACAAACATTTTC--CAATAGTGAGCAAAG
SEQ1302	GGT-TATCTAAAAATCAGTG-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1303	GGT-TATCTAAAAATCAGTG-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1304	GGG-TATCTAAAAA-----
SEQ1305	GCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGGC
SEQ1306	-----
SEQ1307	GCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGGC
SEQ1308	GCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGAC
SEQ1309	GGT-TATCTAAAAATCAGTA-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1310	GGT-TATCTAAAAATCAGTA-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1311	GCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGAC
SEQ1312	GCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGGC
SEQ1313	GGT-TATCTAAAAATCAGTG-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1314	GCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGAC
SEQ1315	GCTCTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTTTTGGAGGGGC
SEQ1316	GGT-TATCTAAAAATCAGTA-CCGGTATTAATGAAAAAGTTCTTGTGGGGGGATTGAA
SEQ1301	TGAATTTTCCATCAATCTTGG--TAATTTTTGAAAAATGTTTCTTTTAGTTTCTAAC
SEQ1302	GTAGTTCTCTTCAACCTATGAGACGTTACGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1303	GTAGTTCTCTTCAACCTATGAGACGTTACGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1304	-----
SEQ1305	TTAGCTTACGGACACCCTTA--TGCCTGCTCAGGAATTATTAATATCC-----
SEQ1306	-----
SEQ1307	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1308	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1309	GTAGTTCTCTTCAACCTATGAGACGTTATGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1310	GTAGTTCTCTTCAACCTATGAGACGTTATGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1311	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1312	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1313	GTAGTTCTCTTCAACCTATGAGACGTTACGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1314	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1315	TTAGCTTACGGACACCCTTA---TGCCTGCTCAGGAATTATTAATATCCTTCATCTTAT
SEQ1316	GTAGTTCTCTTCAACCTATGAGACGTTATGCTAAAGAAGATAATCGTAACGGAGAATAT
SEQ1301	CCTTGATCTCGCATCCCTTCCATTGGTAAGATTACYTCTTCTAAATAGCCACCTTGTTT
SEQ1302	CCGTTGCTCAGTTTTCTCCTGACTCTTATGCTG--AAACTGTAATGTTAGAAGGGGCAC
SEQ1303	CCGTTGCTCAGTTTTCTCCTGACTCTTAKGCTG--AAACTGTAATGTTAGAAGGGGCAC
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	CAGGCATTAAAAATATAAAAAATAAACCTATGGGC-CTAACTGCCATTGCAGGGGCA----
SEQ1308	CAGGCATTAAAAATATAAAAAATAAACCTATGGGT-CTAACTGCCATTGCAGGGGCAG---
SEQ1309	CAGTTGCTCAGTTTTCTCCTGACTCTTATGCTG--AAACTGTAATGTTAGAAGGGGCCC
SEQ1310	CAGTTGCTCAGTTTTCTCCTGACTCTTATGCTG--AAACTGTAATGTTAGAAGGGGCCC
SEQ1311	CAGGCATTAAAAATATAAAAAATAAACCTATGGGT-CTAACTGCCATTGCAGGGGCAGGA-
SEQ1312	CAGGCATTAAAAATATAAAAAATAAACCTATGGGTTCTAACTGC-----
SEQ1313	CCGTTGCTCAGTTTTCTCCTGACTCTTATGCTG--AAACTGTAATGTTAGA-----
SEQ1314	CAGGCATTAAAAATATAAAAAATAAACCTATGGGT-CTAACTGCCATTGCAGGGGC----
SEQ1315	CAGGCATTAAAAATATAAAAAATAAACCTATGGGT-CTAACTGCCATTGCAGGGGCAGGA-
SEQ1316	TABCMARATVSTNCSRATNGTSAGTHAS-----

Table 13: Comparative Sequences relating to SAG0466 (S. glase)

SEQ1301	GCTGTTAAGGCGCGTTTATGGCTCAAGAATGCCAATTTATCTAACATTTCTCTTCTAAA
SEQ1302	AAGAGTCTGTCAAAAATATGGTTTTAGAAAGAGAAATGTTAGATAAAATTGGCATTCTTGA
SEQ1303	AAGAGTCTGTCAAAAATATGGTTTTAGAAAGAGAAATGTTAGATAAAATTGGCATTCTTGA
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	GAGAGTCTGTCAAAAATATGGTTTTAGAAAGAGAAATGTTAGATAAAATTGGCATTCTTGA
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	CCATATTTTTGACAGACTCTCTGGGCCCCCTT--CTAACATTACAGTTTCAGCATAAGAG
SEQ1302	CCATAAACGCGCCTTAACAGCTAAACAAGGTGGCTATTTAGAAAGAGGTAATCTTACCAA
SEQ1303	CCATAAACGCGCCTTAACAGCTAAACAAGGTGGCTATTTAGAAAGAGGTAATCTTACCAA
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	CCATAAACGCGCCTTAACAGCTAAACA-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	CAGGAGAAAACCTGAGCAACTGTATATTCTCCGTTACGATTATCTTCTTTAGCATAACGT
SEQ1302	GGAAGGGATGCGAGATCAAGGCGTTAGAAAACATAAAGAAGCATTTTTTTCAAAAATTAC
SEQ1303	GGAAGGGATGCGAGATCAAGGCGTTAGAAAACATAAAGAAGCATTTTTTTCAAAAATTAC
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----

S. 94116.5 (3112/112)

Table 13: Comparative Sequences relating to SAG0466 (base)

SEQ1301	TCATAGGTTGAAGAGAACTACTTTCAATCCCCCAACAAGAACTTTTCATTAATACCG
SEQ1302	AAGATTGATGGAAAATTACCTTTGCTCACTATTGGAAATGTTGTTTAATGCACGATG
SEQ1303	AAGATTGATGGRAAATTACCTTTGCTCACTATTGGAAATGTTGTTTAATGCACGATG
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	TACTGATTTTTAGATAACCAAAAAAC--AAGGCAGAACTTGATGAAGCACACTGCATAT
SEQ1302	TGCTGCATTTCTAACGCTTCAGAGTCAGAAAACAGAATTCCGGATTGTTTCACATTGTAG
SEQ1303	TGCTGCATTTCTWACGCTTCAGAGTCAGAAAACAGAATTCCGGATTGTTTCACATTGTAG
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	AATCGTTTGTACTGGAATATAGGATTCATAATCAGAAAAAGAGTCATCAAACGACCAA
SEQ1302	AGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAACTAT
SEQ1303	AGTAGCAGGAGATCCCAAGCTTAGTCCAGAATTGGTTCACACGGCTACGGAAAACTAT
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----

Table 13: Comparative Sequences relating to SAG0466 (polase)

SEQ1301	ATTGCCCCCAGTACCAACTGTGTTCCACAAATAATACTATCAATGTTAGATTCTGATT
SEQ1302	AACAGAAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTG
SEQ1303	AACAGAAACTCATACTAAAATATCGGATTATGATGCAATTGAATGGAATGAACCATTG
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	TATTTTTTTTATTTGATTTAAAAGGTGTGCTCCTAAAAGTTCTGGACGGTAAGTTTAAA
SEQ1302	AGCGATTGATGCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTT
SEQ1303	AGCGATTGATGCTTTATTTAATCATTATTATCCTGAAGAGAGAGAAAAATTCAATATTT
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----
SEQ1301	TGCTT-----
SEQ1302	TGGAGGGGCATTAGCTTACGGACACCCTTATGCCTGCTCAGGAATTA
SEQ1303	TGGAGGGGCATTAGCTTACGGACACCCTTATGCCTGCTCAGG-----
SEQ1304	-----
SEQ1305	-----
SEQ1306	-----
SEQ1307	-----
SEQ1308	-----
SEQ1309	-----
SEQ1310	-----
SEQ1311	-----
SEQ1312	-----
SEQ1313	-----
SEQ1314	-----
SEQ1315	-----
SEQ1316	-----

Table 14: Comparative Sequences relating to SAG0471 (gluc kinase)

SEQ ID NO. 1401: SAG0471 FROM THE 18RS21 GBS TYPE II STRAIN

SEQ ID NO. 1401: SAGUO1 FROM THE 18RS21 GBS TIME 11:03AM
TTAAATTTGGTATCTTGACGCTTGAGGGAAGAAGTACAAGAAAAATGGGCAATTGAGACCAATACTTTAGAAAAACGGAAGACATATCGTT
TCTGATATCGCTTGAATCTCTCAAACATCGTTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGGTATCGGTATGGGTTCTCCAGG
AGCTGTTGATAGAAGTAGTAAACAGTAACAGGTGCTTTTAATCTAAATTTGGGCTGATACCTCAAGAAGTAGGTTCACTTATTGAAAAG
AAGTTGGAATTCCATTTTTTATTGTAACAGTAGTCTAATGTTCGAGCACTTGGTGAACGCTGGGTAGGTGCTGGTGGCCAATAAATCCCGAC
GTTGTTTTCTCGTAACCTTCGGAACAGAGTAGGTGGAAGTGTTATCGAGATGGTAACCTCATCCATGGTGGTTGCAGGAGCAGGTGGAGA
AATTGGGCTCATGATTGTTGATCCAGAAAATGGATTACGTGCACATGTGGTAACAAAGGCTGCCTTGAGACAGTTGCATCAGCGACAG
GTGTTGTTAGAGTAGCACGTCAACTCGCAGAACAATATGAGGGTTCGTCTGCCATTAAAGCAGCGATTGACACCGGTGATACGTGTTACA
AGTAAAGATATTTTTATAGCAGCAAGATGGGGGATAAAATTTGCTAATTCGTGTTGTTGAACGTGTATCACTTTACCTTGGACTGGCAGC
AGCTAATATTTCAAATATTTTAAACCTTGATTCGTGGTTATTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTTACGTAGTCGCGTTG
AGAAATACTTTGTGCATATTTGCTTTCCCAAGTTAAAAAGTCACTAAAATTAAGAT

SEQ ID NO. 1402: SAG0471 FROM THE 090 GBS TYPE Ia STRAIN

SEQ ID NO. 1402: SAG0471 FROM THE U90 GBS TYPE 1A STRAIN
CGTTTCTGATATCGTTGAATCTCTCAACACATCGTTTGGAGCCTCTATGGATTAAACAAAGATGACTTTCTCGGTATCGGTATGGGTTCTC
CAGGAGCTGTTGATAGAAGTCTAGTAAACAGTAACAGGTGCTTTTAATCTAAATTGGGCTGATACTCAAGAAAGTAGGTTTCGGTTATTGAA
AAAGAAGTTTGAATTCATTTTTTATTGATAACGATGCTAATGTTGCAGCACATTGGTGAACCTCGGTTAGGTGCTGGTGCCCAATAATCC
CGATGTTGTTTTCGTAACCCCTCGGAACAGGAGTAGGTGGAGGTTGATATCGCAGATGGTAACCTCATCCATGGTGTTCGAGGACAGGTG
GAGAAATTTGGGCATATGATTGTTGATCCAGAKAATGGATTTACGTTGCACATGTGGTAACAAAGGCTGTCTTGAGACAGTTGCATCAGCG
ACAGGTGTTGTTAGAGTAGCACGTCAACTCGCAGAACATATGAAGGTTTCGTCTGCCATTAAAGCAGCGATTGACAACGGTGATACTGT
TACAAGTAAAGATATTTTTATAGCAGCAGAAGATGGGGATAAATTTGCTAATTCGTGTGTTGAACGTGTATCAGGTTACCTTTGGACTGG
CAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCCTGTGGTTATTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGC
GTTTGAGAAATCTTTTGTCACATTTG

SEQ ID NO. 1403: SAG0471 FROM THE COH1 GBS TYPE Ia STRAIN

SEQ ID NO. 1403: SAG0471 FROM THE COHI GES TYPE 1a STRAIN
ACAAGAAAAATGGGCAATTGAGACCAATACTTTAGAAAAACGGAAGACATATCGTTTCTGATATCGTTGAATCTCTCAAACATCGTTTGA
GCCTCTATGGATTAAACAAAGATGACTTTCTCGGTATPCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAACAGTAACAGGT
GCTTTTAACTCAAATTTGGGCTGATACTCAAGA

SEQ ID NO. 1404: SAG0471 FROM THE CJB110 GBS NONTYPEABLE STRAIN

SEQ ID NO. 1404: SAG0471 FROM THE CJE110 GBS NONIFERABLE STRAIN
TTGGTATCTTGACGCTTGAGGAGAAGTACAAGAAAAATGGGCAATTGAGACCAATACTTTAGAAAAACGGAAGACATATCGTTTCTGATA
TCGTTGAATCTCTCAACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGGTATCGGTATGGGGTCTCCAGGAGCTGTT
GATAGAACTAGTAAAC

SEO ID NO. 1405: SAG0471 FROM THE CJB110 GBS NONTYPEABLE STRAIN

SEQ ID NO. 1405: SAG0471 FROM THE CJB110 GBS NONTYPEABLE STRAIN
CACCAGCTAATATTTCAAATATTTTAAACCCGTATTCTGTGTTTATTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTTACGTAGTCGC
GTTGAGAAATACTTTTGTCAATTTTGTCTTTCCCACTGTAAAAAGTCAACTA

SEO ID NO. 1406: SAG0471 FROM THE 2603V/R GBS TYPE V STRAIN

SEQ ID NO. 1406: SAG0471 FROM THE 2803V/R GBS 1125 V. JAMES
GGGCAATTGAGACCAATACTTTAGAAAACGGGAAGACATATCGTTTCTGATATCGTTGAATCTCTCAAACATCGTTTGAGCCTCTATGGA
TTAAACAAAAGATGACTTTCTCGGTATCGGTATGGGTCTCCAGGAGCTG

SEO ID NO. 1407: SAG0471 FROM THE H36b GBS TYPE Ib STRAIN

SEQ ID NO. 1407: SAG0471 FROM THE H365 GBS TYPE 1D STRAIN
GGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTCTGATATCGTTGAATCTCTCAAACATCGTTTGAGCCTCTATGGAT
TAACAAAAGATGACTTTCTCGGTATCGGTATGGTTCTCCAGGACTGTTGTATGAGAAGTAGTAAAAACAGTAACAGGTGCTTTTAATCTA
AATTGGGCTGATACCTTAAGCAGTAGGTTTCAGTTATTGAAAAAGAAGTTGGAATTCCATTTTTTATTGATAACGATGCTAATGTGTCAGC
ACTTGGTGAACGCTGGGTAGGTGCTGGTGCCAATAATCCCACGTTGTTTTTCGTAACC

SEQ ID NO. 1408: SAG0471 FROM THE H36 GBS TYPE 1b STRAIN (REVERSE COMPLEMENT)

SEQ ID NO. 1408: SAG0471 FROM THE H36 GBS TYPE 1b STRAIN (REVERSE COMPLEMENT),
GAGACAGTTGCATCAGCGACAGGTGTTGTTAGAGTAGCACGTCACATCGCAGAACATATGAGGGTTCGCTCGCCATTAAAGCAGCGAT
TGACAACCGGTGATAGTCTGTTACAAGTAAAGATATTTTATAGCAGCAGAAGATGGGGATAAAATTTGCTAATTCTGTTGTTGAACGTGTAT
CAGCTTACCTTGGAGCTGGCAGCAGCTAATATTTTCAAAATATTTTAAACCTTGATTCTGTGGTTATTGGTGGCGGTGTCTCAGCAGCAGGT
GAATTTTACGTAGTCGCGTTGAGAAATACTTTGTCAATTTGCTTTCCACA

SEO ID NO. 1409: SAG0471 FROM THE M732 GBS TYPE III STRAIN

SEQ ID NO: 1409: SAG0471 FROM THE M/32 GBS TYPE III STRAIN
ACAAGAAAAATGGCGAATTGAGACCTACTTAGAAAAACGGAAGACATATCGTTTCTGATATCGTTGAATCTCTCAAACATCGTTTGAGC
CTCTATGGGATTAACAAAAGATGACTTTCTCGGTATCGGTATGGGTCTCCAGGAGCTGTTGATAGAAGTAGTAAACAGTAACAGGTGC
TTTAAATCTAAATTGGGCTGATACTCAAGAAGTAGGTTTCGGTTATTGAAAAGAAGTTGGAAATCCAACTTTTATTGATAACGATGCTA
ATGTTGCAGCACTTGGTGAACGCTGGGTAGGTGCTGGTGCACCAATACCTCCCGATTTGTGTTTTTCGTAACCCCTCGGAACAGGATAGGTGGA
GTGTTTATCGCAGATGGTAACCTCATCATCGGTGTTGCAAGAGCAGGTGGAGAATTGGGCATATGATT

SEQ ID NO. 1410: SAG0471 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

SEQ ID NO. 1410: SAG0471 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
CAGCAGCAGGTGAATTTTACGTCAGTAGCGCTTGAGAATACTTTGTCACATTGGTTTTCCCAAGTTAAAAAGTCAACTAAAATTAAG
ATTGCTGAAGTAAAGT

SEQ ID NO. 1411: SAG0471 FROM THE M781 GBS TYPE III STRAIN

SEQ ID NO. 1411: SAG0471 FROM THE M781 GBS TYPE III STRAIN
AGAAGTACAAGAAAATGGGCAATTGAGACCATACTTAGAAAAAGGAAGACATATCGGTTTCTGATATCGTTGAATCTCTCAAACATCGTT
TGAGCCTCTTATGGATTACAAAAGATGACTCTTCGGTATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAAGTAGTAAACAGTAACA
GGTGCTTTTAACTCTAAATGGGGCTGATACTCAAGAAGTAGGTTTCGGTTATTGAAAAAGAGTTGGAATTCCATTTTTTATTGATAACGA
TGCTAATGTTGCAGCACTTGGTGAACGCTGGGTAGGTGCTGGTGCCAATAATCCCGATGTTGTTTTCGTAACCTCGGAACAGGAGTA

உள்ளுறை

SEQ ID NO. 1412: SAG0471 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
GATACTGTTACAAGTAAAGATATTTTTATAGCAGCAGAAGATGGGGATAAATTGCTAATTCTGTTGTTGAACGTGTATCACGTTACCT
TGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTAC
GTAGTCGCGTTGAGAAATACTTTGTCACATTTGCTTTCCACAAGTTAAAAA

SEQ ID NO. 1413: SAG0471 FROM THE 090 GBS TYPE Ia STRAIN
AAATTGGTATCTTTGAGCCTTGAGGAGGAAGTACAAGAAAATGGGCATTGAGACCTACTTAGAAAAACGGAAGACATATCGTTTCTGA
TATCGTTGAATCTCTACCAACATCGTTTGAGCCTCTATTGGATTAAACAAAAGATGACTTTCTCGGTATCGGTATGGGTTCTCCAGGAGCTG
TTGATAGAAGTCTAGTAAACAGTAACAGGTGCTTTTAATCTAAATTGGGCTGATACTCAAGAAGTAGGTTTCAGTTATTGAAAAGAAGTT
GGAATTCCTATTTTTTATGTATAACGATGCTAATGTTTCAGCACTTGGTGAACGCTGGGTAGGTGCTGGTGCCATAATCCCGACGTTGT
TTTCGTAAACCTCGGAACAGGAGTAGTGGAGG

SEQ ID NO. 1414: SAG0471 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)
GTGATACTGTTACAAGTAAAGATATTTTATAGCAGCAGAAGATGGGGATAAATTTGCTAATTCGTGTTGAACGTGTATCAGGTTAC
CTTGGACTGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTT
ACGTAGTCGCGTTGAGAAATACTTTATCACATTTGCTTTCCACAAGTTAAAAAGTCAACTAAAATTAAGATTG

SEQ ID NO. 1415: SAG0471 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)
GTTATCGCAGATGGTAACCTCATCCATGGTGTTCAGGAGCAGGTTGGAGAAATTTGGCGCATATATGTTGTATCCAGAAAATGGATTAC
GTGCACATGTGGTAAACAAGGCTGCCTTGAGACAGTTGCATCAGCGACAGGTGTTGTTAGAGTAGCAGCTCAACTCGCAGAAACAATATG
AGGGTTCGTCTGCCATTAAAGCAGCGATTGACCACAGGTGATACGTTTACAAGTAAAGATATTTTTATAGCAGCAGAAGATGGGGATAAA
TTTGCTAATTCTGTTGTTGAACGTGTATCACGTTACCTTGGACTGCGCAGCAGCTAATATTTCAAATATTTTAAACCTTGATTCTGTGGT
TATTTGGTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTTGTACATTGTTGCTTCCCAAGTTAAAA
AGTCAAGTAA

SEQ ID NO. 1416: SAG0471 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)
TGGTATCTTGACGCTTGAGGGAGAAGTACAAGAAAAATGGGCAATTGAGACCATACTTAGAAAACGGAAGACATATCGTTTCTGATATC
GTTGAATCTCTCAAACATCGTTTGAGCCTCTATGGATTAAACAAAGATGACTTTCTCGGTATCGGTATGGGTTCTCCAGGAGCTGTTGA
TAGAACTAGTAAACAGTCACAGGTGCTTTTAATCTAAATTGGGCTGATACTCAAGAAGTAGGTTCAGTTATTGAAAAAGAGCTGGAA
TTCCATTTTTTATTG

SEQ ID NO. 1417: SAG0471 FROM THE 2603V/R TYPE V GBS STRAIN (REVERSE COMPLEMENT)
AGCAGCTAATATTTCAAAATATTTTAAACCTGATTCTGTGGTTATTGGTGGCGGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCG
TTGAGAAATACTTTGTACATTTGTTTTCCACAAGGT

```

SEQ1401_
SEQ1402
SEQ1403
SEQ1404
SEQ1405
SEQ1406
SEQ1407
SEQ1408
SEQ1409
SEQ1410
SEQ1411
SEQ1412
SEQ1413
SEQ1414
SEQ1415
SEQ1416
SEQ1417

```

```

SEQ1401_
SEQ1402
SEQ1403
SEQ1404
SEQ1405
SEQ1406
SEQ1407
SEQ1408
SEQ1409
SEQ1410
SEQ1411
SEQ1412
SEQ1413
SEQ1414
SEQ1415
SEQ1416
SEQ1417

```

Table 14: C mparative Sequences relating to SAG0471 (gluc kinase)

SEQ1401	-----
SEQ1402	-----
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	CAGTTGCATCAGCGACAGGTGTTGTTAGAGTAGCACGTCAACTCGCAGAACAAATATGAG
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	CAGTTGCATCAGCGACAGGTGTTGTTAGAGTAGCACGTCAACTCGCAGAACAAATATGAG
SEQ1416	-----
SEQ1417	-----
SEQ1401	-----
SEQ1402	-----
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	GTTTCGTCTGCCATTAAAGCAGCGATTGACAACGGTGATACTGTTACAAGTAAAGATATT
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----GATACTGTTACAAGTAAAGATATT
SEQ1413	-----
SEQ1414	-----GTGATACTGTTACAAGTAAAGATATT
SEQ1415	GTTTCGTCTGCCATTAAAGCAGCGATTGACCACGGTGATACTGTTACAAGTAAAGATATT
SEQ1416	-----
SEQ1417	-----
SEQ1401	-----TTAAATTTGGTATCTTGACGCTTGAGGGAGAAGTACAA
SEQ1402	-----
SEQ1403	-----ACAA
SEQ1404	-----TTGGTATCTTGACGCTTGAGG-AGAAGTACAA
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	TTATAGCAGCAGAAGATGGGGATAAATTTGCTAATCTGTTGTTGAACGTGTATCACGT
SEQ1409	-----ACAA
SEQ1410	-----
SEQ1411	-----AGAAGTACAA
SEQ1412	TTATAGCAGCAGAAGATGGGGATAAATTTGCTAATCTGTTGTTGAACGTGTATCACGT
SEQ1413	-----AAATTTGGTATCTTGACGCTTGAGGGAGAAGTACAA
SEQ1414	TTATAGCAGCAGAAGATGGGGATAAATTTGCTAATCTGTTGTTGAACGTGTATCACGT
SEQ1415	TTATAGCAGCAGAAGATGGGGATAAATTTGCTAATCTGTTGTTGAACGTGTATCACGT
SEQ1416	-----TGGTATCTTGACGCTTGAGGGAGAAGTACAA
SEQ1417	-----

Table 14: Comparative Sequences relating to SAG0471 (glucokinase)

SEQ1401	AAAAATGGGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1402	-----CGTTTCTGATATC
SEQ1403	AAAAATGGGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1404	AAAAATGGGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1405	-----CACCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1406	-----GGGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1407	-----GGCAATTGAGACCAATACTTTAGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1408	ACCTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1409	AAAAATGGGCAATTGAGACCA-TACTT-AGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1410	-----
SEQ1411	AAAA-TGGGCAATTGAGACCA-TACTT-AGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1412	ACCTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1413	AAAAATGGGCA-TTGAGACCA-TACTT-AGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1414	ACCTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1415	ACCTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1416	AAAAATGGGCAATTGAGACCA-TACTT-AGAAAACGGAAGACATATCGTTTCTGATATC
SEQ1417	-----AGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATT
SEQ1401	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1402	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1403	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1404	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1405	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1406	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1407	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1408	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1409	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1410	-----CAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1411	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1412	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1413	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1414	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1415	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1416	TTGAATCTCTCA-AACATCGTTTGAGCCTCTATGGATTAACAAAAGATGACTTTCTCGG
SEQ1417	GTGGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTT---
SEQ1401	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1402	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1403	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1404	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1405	GTCACATTTGCTTTCCACAAGTTAAAAAGTCAACTA-----
SEQ1406	ATCGGTATGGGTTCTCCAGGAGCTG-----
SEQ1407	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1408	GTCACATTTGCTTTCCACA-----
SEQ1409	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1410	GTCACATTTGCTTTCCACAAGTTAAAAAGTCAACTAAAATTAAGATTGCTGAAGTAGG
SEQ1411	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1412	GTCACATTTGCTTTCCACAAGTTAAAAAGTCAACTAAAATTAAGATTG-----
SEQ1413	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTAACAGGTGCTTT
SEQ1414	ATCACATTTGCTTTCCACAAGTTAAAAAGTCAACTAAAATTAAGATTG-----
SEQ1415	GTCACATTTGCTTTCCACAAGTTAAAAAGTCAACTAA-----
SEQ1416	ATCGGTATGGGTTCTCCAGGAGCTGTTGATAGAACTAGTAAAACAGTCACAGGTGCTTT
SEQ1417	GTCACATTTGTTTCCACAAGGT-----

SEQ1401	ATCGCAGATGGTAACCTCATCCATGGTGTTCAGGAGCAGGTGGAGAAATTGGGCATAT
SEQ1402	ATCGCAGATGGTAACCTCATCCATGGTGTTCAGGAGCAGGTGGAGAAATTGGGCATAT
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	ATCGCAGATGGTAACCTCATCCATGGTGTTCAGGAGCAGGTGGAGAAATTGGGCATAT
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----
SEQ1401	ATTGTTGATCCAGAAAATGGATTTACGTGCACATGTGGTAACAAAGGCTGCCTTGAGAC
SEQ1402	ATTGTTGATCCAGAKAATGGATTTACGTGCACATGTGGTAACAAAGGCTGTCTTGAGAC
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	ATT-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----
SEQ1401	GTTGCATCAGCGACAGGTGTTGTTAGAGTAGCACGTCAACTCGCAGAACAAATATGAGGG
SEQ1402	GTTGCATCAGCGACAGGTGTTGTTAGAGTAGCACGTCAACTCGCAGAACAAATATGAAGG
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----

Table 14: C mparative Sequences relating to SAG0471 (gluc kinase)

SEQ1401	TCGTCTGCCATTAAAGCAGCGATTGACACCGGTGATACTGTTACAAGTAAAGATATTTT
SEQ1402	TCGTCTGCCATTAAAGCAGCGATTGACAACCGGTGATACTGTTACAAGTAAAGATATTTT
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----
SEQ1401	ATAGCAGCAGAAGATGGGGATAAATTTGCTAATTCTGTTGTTGAACGTGTATCACGTTA
SEQ1402	ATAGCAGCAGAAGATGGGGATAAATTTGCTAATTCTGTTGTTGAACGTGTATCACGTTA
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----
SEQ1401	CTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATTGG
SEQ1402	CTTGGACTGGCAGCAGCTAATATTTCAAATATTTTAAACCCTGATTCTGTGGTTATTGG
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----

Table 14: Comparativ Sequences relating t SAG0471 (gluc kinase)

SEQ1401	GGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTTGTCAC
SEQ1402	GGCGGTGTCTCAGCAGCAGGTGAATTTTACGTAGTCGCGTTGAGAAATACTTTGTCAC
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----
SEQ1401	TTTGCTTTCCCAAGTTAAAAAGTCAACTAAAATTAAGAT
SEQ1402	TTTG-----
SEQ1403	-----
SEQ1404	-----
SEQ1405	-----
SEQ1406	-----
SEQ1407	-----
SEQ1408	-----
SEQ1409	-----
SEQ1410	-----
SEQ1411	-----
SEQ1412	-----
SEQ1413	-----
SEQ1414	-----
SEQ1415	-----
SEQ1416	-----
SEQ1417	-----

Table 15: C mparative Sequences relating to SAG0492

SEQ ID NO. 1501: SAG0492 FROM THE 1169NT1 GBS NONTYPEABLE STRAIN

TGACTTGGATATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCTCT
TGGAGTACCAACAAAGGGAACAGTGACTTTTGAAGGAATTGATATAACAGACAAAAAATGATATTTTAAATGCGCGAAAAAATG
GGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGACTTTC
TAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTG
GAGGACAACAACACGGATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTATGAACCTACTTCAGCTCTTGAT
CCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAGCTAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTT
TGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGATGCAGGCATTATTGTGAGCAAGGGACCCCTAAGGAAGTAT

SEQ ID NO. 1502: SAG0492 FROM THE 18RS21 GBS TYPE II STRAIN

TTGGGAAAAATGAGGTTTAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAG
TCAACATTTTAAAGAACAATGAATCTCTTGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAAGAA
TGATATTTTAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTT
TATCACCTATTAAGACAAAGGGGCTTTCTAATCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAG
GCTAATACCTTATCCAGCTAGCTTATCTGGAGGACAACAACGAATTTGCTATTGCAAGAGGTCTTGCAATGAATCCTCATGTCTCTTCT
TTTTGATGAACCTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAGCTAAATCTGGTATGACGA
TGTTATTGTCACTCATGAAATGGGTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGACGAGAAATTAT

SEQ ID NO. 1503: SAG0492 FROM THE 2603V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)

AAAAATGAGGTTTAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAAC
ATTTTAAAGAACAATGAATCTCTTGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAAGAATGATA
TTTTTAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCA
CCTATTAAGACAAAGGGGCTTTCTAATCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAA
TACTTATCCAGCTAGCTTATCTTGAGGACAACAACGAATTTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTG
ATGAACCTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAGCTAAATCTGGTATGACGATGGTT
ATTGTCACTCATGAAATGGGTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGATGCAGGAATTATTGTTGAGCAAGGGGCC

SEQ ID NO. 1504: SAG0492 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GAGGTTTTAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTT
AAGAACAATGAATCTCTTGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAAGAATGATATTTT
AAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATT
AAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTA
TCCAGCAAGCTTATCTGGAGGACAACAACGAGGATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
CTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAGCTAAATCTGGTATGACGATGGTTATTGTC
ACTCATGAAATGGGTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGATGCAGGATTATTGTTGAGCAAGGGACCCCTAAGAA
AGTAT

SEQ ID NO. 1505: SAG0492 FROM THE 090 GBS TYPE Ia STRAIN

TGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCTCTTGAAGTACCAACAAAGGGAACAGTG
ACTTTTGAAGGGATTGATATAACAGACAAAAAGAATGATATTTTAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCT
ATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCAT
ACGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTGGAGGGCAACAACGAATTTGCTATT
GCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAACCTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGAC
TGTTATGCAAGATTTAGCTAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTGCACGTGAAGTAGCGGATCGTGTCA
TTTTTATGGATGCAGGCATTATTGTTgAsCAAGGGACCCCTAAGGAAGTA

SEQ ID NO. 1506: SAG0492 FROM THE A909 GBS TYPE Ia STRAIN

CAATACAAGGACTTCATAAAGTTTTGGGAAAAATGAGGTTTTAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTAGTGGTTATT
ATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCTCTTGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGG
GATTGATATAACAGACAAAAAGAATGATATTTTAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATA
TGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTT
GAAAAAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAACGAATTTGCTATTGCAAGAGGTCT
TGCAATGAATCCTGATGTCCTTCTTTTTGATGAACCTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAG
ATTTAGCTAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGAT
GCAGGAATTATTGTgAGCAAGGGGCCCTAAGGAAGTATTGAGCAGACAAAAGAAATCCGCACAGAGATTCTT

SEQ ID NO. 1507: SAG0492 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

GACTTGGATATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCTCTT
GGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAAGAATGATATTTTAAATGCGCGAAAAAATGG
GCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGGACTTCT
AAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTGG
AGGACAACAACGAATTTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAACCTACTTCAGCTCTTGATC
CTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAGCTAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTT
GCACGTGAAGTAGCGGATCGTGTCTTTTTTATGGATGCGGGAATTATTGTGAGCAAGGGACC

Table 15: Comparative Sequences relating to SAG0492

SEQ ID NO. 1508: SAG0492 FROM THE H36b GBS TYPE Ib STRAIN

ATGAGGTTTTAAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTT
TTAAGAACAAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAGAATGATATTTT
TAAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
TTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATACT
TATCCAGCTAGCTTATCTGGAGGACAACAACAACGAATTGCTATTGCAAGAGGCTTTGCAATGAATCCTGATGTCCTTCTTTTGATGA
ACCTACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTAGCTAAATCTGGTATGACGATGGTTATTG
TCACTCATGAAATGGGTTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGATGCASGAATTATTGTTGAGCAAGGGGCCCTAAG
GAAGTAT

SEQ ID NO. 1509: SAG0492 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)

GGTTTTAAAAGGCATTGACTTGGATATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAA
GAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAGAATGATATTTTAA
ATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTA
GACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATACTTATC
CAGCTAGCTTATCTGGAGGACAACAACAACGAATTGCTATTGCAAGAGGCTTTGCAATGAATCCTGATGTCCTTCTTTTGATGAACCT
ACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTAGCTAAATCTGGTATGACGATGGTTATTGTCAC
TCATGAAATGGGTTTTGCACGTGAAGTAGCGGATCGTGTCAATTTTATGGATGCAGGAATTATTGTTGAGCAAGGGGCCCTAAGGAAG
TATTTAGCAAAACAAAGAAAT

SEQ ID NO. 1510: SAG0492 FROM THE M732 GBS TYPE III STRAIN

GGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGA
CTTTTGAAGGGATTGATATAACAGACAAAAGAATGATATTTTAAATGCGCGAAAAAATGGGCATGGTTTTTCAACAGTTCAATCTA
TTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATA
CGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCAAGCTTATCTGG

SEQ ID NO. 1511: SAG0492 FROM THE COH1 GBS TYPE Ia STRAIN

ATTGACTTGGATATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACATTTTAAAGAACAATGAATCT
CTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGATTGATATAACAGACAAAAGAATGATATTTTAAATGCGCGAAAAA
TGGGCATGGTTTTTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTATTAAGACAAAGGGACTT
TCTAAGCTTGATGCTCAGACAAAAGCATAACGAGCTACTTGAAAAAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCAAGCTTATC
TGG

Table 15: C mparativ Sequences relating to SAG0492

SEQ1501	-----TGACTTGG
SEQ1502	-----TTGGGAAAAATGAGGTTTTAAAAGGCATTGACTTGG
SEQ1503	-----AAAAATGAGGTTTTAAAAGGCATTGACTTGG
SEQ1504	-----GAGGTTTTAAAAGGCATTGACTTGG
SEQ1505	-----
SEQ1506	AATACAAGGACTTCATAAAAGTTTTGGGAAAAATGAGGTTTTAAAAGGCATTGACTTGG
SEQ1507	-----GACTTGG
SEQ1508	-----ATGAGGTTTTAAAAGGCATTGACTTGG
SEQ1509	-----GGTTTTAAAAGGCATTGACTTGG
SEQ1510	-----
SEQ1511	-----ATTGACTTGG
SEQ1501	TATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1502	TATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1503	TATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1504	TATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1505	-----TGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1506	TATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1507	TATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1508	TATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1509	TATTCATCAAGGAGAAGTAGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1510	-----GGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1511	TATTCATCAAGGAGAAGTGGTGGTTATTATTGGCCCTTCTGGCTCTGGTAAGTCAACAT
SEQ1501	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1502	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1503	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1504	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1505	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1506	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1507	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1508	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1509	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1510	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1511	TTTTAAGAACAATGAATCTCTTGGAAGTACCAACAAAGGGAACAGTGACTTTTGAAGGGA
SEQ1501	TTGATATAACAGACAAAAAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1502	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1503	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1504	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1505	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1506	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1507	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1508	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1509	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1510	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1511	TTGATATAACAGACAAAAAGAATGATATTTTTAAATGCGCGAAAAATGGGCATGGTTT
SEQ1501	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1502	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1503	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1504	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1505	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1506	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1507	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1508	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1509	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1510	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA
SEQ1511	TTCAACAGTTCAATCTATTTCCCAATATGACTGTACTAGAAAATATTACTTTATCACCTA

Table 15: Comparative Sequences relating to SAG0492

SEQ1501	TTAAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAA
SEQ1502	TTAAGACAAAGGGGCTTTCTAATCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1503	TTAAGACAAAGGGGCTTTCTAATCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1504	TTAAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAA
SEQ1505	TTAAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAA
SEQ1506	TTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1507	TTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1508	TTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1509	TTAAGACAAAGGGGCTTTCTAAGCTTGATGCTCAGACAAAAGCATATGAGCTACTTGAAA
SEQ1510	TTAAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAA
SEQ1511	TTAAGACAAAGGGGACTTTCTAAGCTTGATGCTCAGACAAAAGCATACGAGCTACTTGAAA
SEQ1501	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1502	AAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1503	AAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1504	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCAAGCTTATCTGGAGGACAACAAC
SEQ1505	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1506	AAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1507	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1508	AAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1509	AAGTTGGACTCAAAGAGAAGGCTAATACTTATCCAGCTAGCTTATCTGGAGGACAACAAC
SEQ1510	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCAAGCTTATCTGG-----
SEQ1511	AAGTTGGACTCAAAGAGAAGGCTAATGCTTATCCAGCAAGCTTATCTGGTABC MARATVS
SEQ1501	ACGGATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1502	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTCATGTCCTTCTTTTTGATGAAC
SEQ1503	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1504	ACGGATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1505	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1506	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1507	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1508	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1509	ACGAATTGCTATTGCAAGAGGTCTTGCAATGAATCCTGATGTCCTTCTTTTTGATGAAC
SEQ1510	-----
SEQ1511	NCSRATNGTSAG-----
SEQ1501	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1502	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1503	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1504	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1505	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1506	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1507	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1508	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1509	TACTTCAGCTCTTGATCCTGAAATGGTAGGTGAAGTCTTGACTGTTATGCAAGATTTAG
SEQ1510	-----
SEQ1511	-----
SEQ1501	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1502	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1503	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1504	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1505	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1506	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1507	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1508	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1509	TAAATCTGGTATGACGATGGTTATTGTCACTCATGAAATGGGTTTTGCACGTGAAGTAG
SEQ1510	-----
SEQ1511	-----

Table 15: C mparative Sequences relating to SAG0492

SEQ1501	GGATCGTGTCAATTTTTATGGATGCAGGCATTATTGT-GAGCAAGGGACCCCTAAGGAAG
SEQ1502	GGATCGTGTCAATTTTTATGGACGCAGAAATTAT-----
SEQ1503	GGATCGTGTCAATTTTTATGGATGCAGGAATTATTGTTGAGCAAGGGGCCC-----
SEQ1504	GGATCGTGTCAATTTTTATGGATGCAGGGATTATTGTTGAGCAAGGGACCCCTAAGAAAG
SEQ1505	GGATCGTGTCAATTTTTATGGATGCAGGCATTATTGTTGASCAAGGGACCCCTAAGGAAG
SEQ1506	GGATCGTGTCAATTTTTATGGATGCAGGAATTATTGTGAGCAAGGGGCCCCTAAGGAAGT
SEQ1507	GGATCGTGTCAATTTTTATGGATGCGGGAATTATTGT-GAGCAAGGGACC-----
SEQ1508	GGATCGTGTCAATTTTTATGGATGCASGAATTATTGTTGAGCAAGGGGCCCCTAAGGAAG
SEQ1509	GGATCGTGTCAATTTTTATGGATGCAGGAATTATTGTTGAGCAAGGGGCCCCTAAGGAAG
SEQ1510	-----
SEQ1511	-----
SEQ1501	AT-----
SEQ1502	-----
SEQ1503	-----
SEQ1504	AT-----
SEQ1505	A-----
SEQ1506	TTTGAGCAGACAAAAGAAATCCGCACAAGAGATTCTT
SEQ1507	-----
SEQ1508	AT-----
SEQ1509	ATTTAGCAAAACAAAAGAAAT-----
SEQ1510	-----
SEQ1511	-----

Table 16: Comparative Sequences relating to SAG0767 (D-alanine - D-alanine ligase)

SEQ ID NO. 1601: SAG0767 FROM THE M781 GBS TYPE III STRAIN

TGGTCGCTCTGTGCGAACGTGAAGTATCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAATTTTTT
GTTAAACTTATTTTATCACGCAAGTAGGTCAATTTATTAACACACAAGAATTTGATGAAATGCCATCTTCAGATGAAA
GTTAATGACAAACCAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTTCC
CCGTTTTACATGGACCAATGGGGGAAGATGGTTCTATCCAAGGATTTTATAGAAAGTTTAAAGGATGCCTTATGTTGGGACT
AATATTCTATCTTCAAGCGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAGGTGTACCTCAGGTTGC
ATATCAAACCTTATTTTGGGGTGTATGATTTGGAACATGCGATTAACTCTCTTTAGAAAACCTTTAAGTTTCCCAATTTTTG
TAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTA
GCTCTCAAGTATGATAGCCGATTTTGTATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAA
TGATGTTAAGACAACCTTTTCTGGCGAAGTTGTTAAAGACGTCGATTTCTATGACTATGACGCCAATATATTGATAATA
AAATTACTATGGATATTCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAA
GCAATCGGGGCTTGTGGTTTATCACGCTGTGATTTCTTTTTGACGAAAGATGGACAAATCTTCTTAAACGAACCTGAATC
AATGCCCGTTTTACTCAGTGGTCAATGTATCCTCTGCTTTGGGAAAATATGGGGCTAACTTATAGTGATTTGATTG

SEQ ID NO. 1602: SAG0767 FROM THE 090 GBS TYPE Ia STRAIN

AAACCGGGCATTTGATTCAGTTCGTTTAAAGAAGACTTGTCCATCTTTTCGTCAAAAAGAAATCACAGCGTGATAAACCACA
AGCCCCGATTGCTTTAAAGCTTTACTTGCATATTGACGATTGCTTCCATAGTTGCTTCATCAACTTTAGCTGGAATAT
CCATAGTAATTTTATTATCAATATATTTGGCGTCATAGTCATAGAAATCGACGCTCTTTAACGACTTCGCCAGGAAAAGTT
GTCTTAACATCATTTATGCTTAAATACCTACTTCAATTTACAGAGCTGTACGCCTTGTTCATCAAAATACGGCTATC
ATACTTGAGAGCTAAGTCAATksCAGAGCGAAGTGAGGATTCATCTGTGCTTTTGAATACCTACTGATGACCCCATAT
TAGCCGGTTTTACAAAATTGGGAACTTAAAGTTTCTAAAGAGAGTTTAAATCGCATGTTCCAAATCATCACCCCTCAAAA
TAAGTTTGATATGCAACCTGAGGTACACCTACTGTTGCAAGGACTTGTGTTTGTGTAATTTTATCCATAGCCACGCTTGA
AGATAGAATATTAGTCCCAACATAAGGCATCCTTAAACTTCTAAAAATCCTTGGATAGAACCATCTTCCCCCATTTGGTC
CATGTAAACGGGGAAAACAATTGCATTATCATCATAGATATCACTTGGACGAACCATTTTGTCTAAATCAACAGTTTGG
TTTGTCTAATACTTTTCTATCTGAAGATGGCATTTTCAATTTCTTGTGTTTAAATAAATTGACCTACTTGGCGT

SEQ ID NO. 1603: SAG0767 FROM THE COH1 TYPE Ia STRAIN

TCGCTCTGCGGAACGTGAAGTATCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAATTTTTTGTTA
AAACTTATTTTATCACGCAAGTAGGTCAATTTATTAACACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
ATGACAAACCAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTTCCCGT
TTTACATGGACCAATGGGGGAAGATGGTTCTATCCAAGGATTTTTATAGAAAGTTTAAAGGATGCCTTATGTTGGGACTAATA
TTCTATCTTCAAGCGTGGCTAT

SEQ ID NO. 1604: SAG0767 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

CGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAAAATTACTATGGATATTCCAGCTAAAGTTGATGAAGCAA
CTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTTCTTT
TTGACGAAAGATGGACAAATCTTCTTAAACGAAGTGAATACAATGCCC

SEQ ID NO. 1605: SAG0767 FROM THE CJB110 GBS NONTYPEABLE STRAIN

AACGTGAAGTATCTGTACTGCTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAATTTTTTGTAAACTTATT
TTATCACGCAAGTAGGTCAATTTATTAACACACAAGAATTTGATGAAATGCCATCTTCAGATGAAA

SEQ ID NO. 1606: SAG0767 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

CTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTAGCTCTCAAG
TATGATAGCCGATTTTTGATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAA
GACAACTTTTCTGGCGAAGTCGTTAAAGACGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAAAATTACTA
TGGATATTTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAGCAATCGGG
GCTTGTGGTTTATCACGCTGTGATTTCTTTTTGACGAAAGATGGACAAATCTTCTTAAACGAAGTGAATACAATGCCGG
TTTTACTCAGTGGTCAATGTATCCTCTGCTTTGGGAAAAT

SEQ ID NO. 1607: SAG0767 FROM THE 18RS21 GBS TYPE II STRAIN (REVERSE COMPLEMENT)

TTGACTTAGCTCTCAAGTATGATAGCCGATTTTGAATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTA
GGCAATAATGATGTTAAGACAACCTTTTCTGGCGAAGTCGTTAAAGACGTCGATTTCTATGACTATGACGCCAAATATAT
TGATAATAAAATTACTATGATATTCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAG
CTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTTCTTTTTGACGAAAGATGGACAAATCTTCTTAAACGAA
CTGAATACAATGCCCGTTTTACTCAGTGGTCAATGTATCCCCCTGCTTTGGGAAAAGTATGGGGCTAACCTT

SEQ ID NO. 1608: SAG0767 FROM THE 18RS21 GBS TYPE II STRAIN

ATCTGTACTGTCTGCAGAAAAGCGTCATGCGTGCTATTAATTATGATAAATTTTTTGTAAACTTATTTTATCACGCAA
GTAGGTCAATTTATTAACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTAATGACAAACCAACTGTTGA
TTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTCCCCGTTTACATGGACCAATGGGGG
AAGATGGTTCTATCCAAGGATTTTATAGAAAGTTTAAAGGATGCCTTATGTTGGGACTAATATCTATCTTCAA

Table 16: Comparative Sequences relating to SAG0767 (D-alanine – D-alanine ligase)

SEQ ID NO. 1609: SAG0767 FROM THE 2603V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)
 GGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAGGTGTACCTCAGGTTGCATATCAAACCTATTTTGGAGG
 GTGATGATTTGGAACATGCGATTAACTCTCTTTAGAACTTTAAGTTTCCCAATTTTTGTAAAACCGGCTAATATGGGG
 TCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCG
 TATTTTGATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTC
 CTGGCGAAGTCGTTAAAGACGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAAAATTACTATGGATATTTCCA
 GCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTT
 ATCAGCGTGTGATTCTTTTGGACGAAAGAATGGACAAATCTTCTTAAACGAACTGAAATAC

SEQ ID NO. 1610: SAG0767 FROM THE 2603V/R GBS TYPE V STRAIN
 TCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAAATTTTTGTAAAACTTATTTTATCACGCAAGT
 AGGTCAATTTATTAAACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTAATGACAAACCAAACCTGTTGATT
 TAGACAAATGGTTCGTCCAAGTGATATCTATGATGATAAT

SEQ ID NO. 1611: SAG0767 FROM THE H36b GBS TYPE 1b STRAIN (REVERSE COMPLEMENT)
 AAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTAG
 CTCTCAAGTATGATAGCCGATTTTGATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAAT
 GATGTTAAGACAACCTTTTCTGGCGAAGTCGTTAAAGACGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAA
 AATTACTATGGATATTTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAG
 CAATCGGGGCTTGTGGTTTATCACGCTGTGATTTCTTTTGGACGAAAGATGGACAAATCTTTTAAACGAACTGAATACA
 ATGCCCGGTTTTACTCAGTGGTCAATGTATCCCTGCTTTGGGAAAATATGGGGCTAACTTATAG

SEQ ID NO. 1612: SAG0767 FROM THE H36b TYPE 1b STRAIN
 CGTGAAGTATCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAAATTTTTGTAAAACTTATTTTAT
 CACGCAAGTAGGTCAATTTATTAAACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTAATGACAAACCAA
 CTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTTCCCGGTTTTACATGGACCA
 ATGGGGGAAGATGGTTCATCCAAGGATTTTGAAGTTTAAAGGATGCCTTATGTTGGGACTAATATTCTATCTTCAAG
 CGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAG

SEQ ID NO. 1613: SAG0767 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
 ATGCGATTAACTCTCTTTAGAACCTTTAAGTTTCCCAATTTTTGTAAACCCGGCTAATATGGGGTCATCAGTAGGTATT
 TCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGATTTTGATTGAACA
 AGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTCTGGCGAAGTTGTTA
 AAGACGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAAAATTACTATGGATATTTCCAGCTAAAGTTGATGAA
 GCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATT
 CTTTTTGGACGAAAGATGGACAAATCTTCTTAAACGAACTGAATACAATGCCCGGTTTTACTCAGTGGTCAATGTATCCTC
 TGCTTTGGGAAAATATGGGGCTAACTT

SEQ ID NO. 1614: SAG0767 FROM THE M732 GBS TYPE III STRAIN
 GTCATGCCGTGCTATTAATTATGATAAAATTTTTGTAAAACTTATTTTATCACGCAAGTAGGTCAATTTATTAAACAC
 AAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTAATGACAAACCAAACCTGTTGATTAGACAAAATGGTTCGTCCA
 AGTGATATCTATGATGATAATGCAATTGTTTTCCCGGTTTTACATGGACCAATGGGGGAAGATGGTTCATCCAAGGATT
 TTTAGAAGTTTAAAGGATGCCTTATGTTGGGACTAATATTCTATCTTCAAGCGTGGCTATGGATAAAATTACAACAAAAC
 AAGTCCTTGCAACAGTAGGTGTACCTCAGG

SEQ ID NO. 1615: SAG0767 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)
 TTTTGAGGGTGATGATTTGGAACATGCGATTAACTCTCTTTAGAACTTTAAGTTTCCCAATTTTTGTAAAACCGGCTA
 ATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCACTTCGCTCTGCAATTGACTTAGCTCTCAAGTAT
 GATAGCCGTATTTTGATTGAACAAGGCGTGACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGAC
 AACTTTTCTGGCGAAGTCGTTAAAGACGTCGATTTCTATGACTATGACGCCAAATATATTGATAATAAAATTACTATGG
 ATATCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCGTCAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCT
 TGTGGTTTATCACGCTGTGATTTCTTTTGGACGAAAGATGGACAAATCTTCTTAAACGAACTGAATACAATGCCCGGTTT
 TACTCAGTGGTCAATGTATCCCCTGCTTTGGGAAAATATGGGGCTAACTTATAGTGA

SEQ ID NO. 1616: SAG0767 FROM THE A909 GBS TYPE Ia STRAIN
 TGGTCGCTCTGCGAAGCTGAAGTATCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTATTAATTATGATAAAATTTTTG
 TAAAACTTATTTTATCACGCAAGTAGGTCAATTTATTAAACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAG
 TTAATGACAAACCAAACCTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTTCCC
 CGTTTTACATGGACCAATGGGGGAAGATGGTTCATCCAAGGATTTTGAAGTTTAAAGGATGCCTTATGTTGGGACTA
 ATATTCTATCTTCAAGCGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAGG

Table 16: Comparative Sequences relating t SAG0767 (D-alanine – D-alanine ligase)

SEQ ID NO. 1617: SAG0767 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)

AAGCAGGGGATACATTGACCACTGAGTAAACCGGGCATTGTATTTCAGTTCGTTTAAGAAGATCTGTCCATCTTTCGTCA
AAAAGAAATCACAGCGTGATAAACCAAGCCCCGATTGCTTTAAAAGCTTTACTTGCAATATTGACGCATTGCTTCCATA
GATGCTTCATCAACTTTAGCTGGAATATCCATAGCAATTTTATTATCAATATATTTGGCG

SEQ1601	GGTCGCTCTGTCTGGAACGTGAAGTATCTGTACTGTCTGCAGAAAGCGTCATGCGTGCTA
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

SEQ1601	TAATTATGATAAAATTTTTTGTAAACTTATTTTATCACGCAAGTAGGTCAATTTATTA
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

SEQ1601	AACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTAATGACAAACCAACTG
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

Table 16: Comparative Sequences relating to SAG0767 (D-alanine - D-alanine ligase)

SEQ1601	TGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATGATGATAATGCAATTGTTTTCC
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	CGTTTTACATGGACCAATGGGGGAAGATGGTTCTATCCAAGGATTTTTAGAAAGTTTTAA
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	GATGCCTTATGTTGGGACTAATATTCTATCTTCAAGCGTGGCTATGGATAAAATTACAA
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----GGCTATGGATAAAATTACAA
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

Table 16: Comparative Sequences relating to SAG0767 (D-alanin - D-alanine ligase)

SEQ1601	AAAACAAGTCCTTGCAACAGTAGGTGTACCTCAGGTTGCATATCAAACCTATTTTGAGG
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	AAAACAAGTCCTTGCAACAGTAGGTGTACCTCAGGTTGCATATCAAACCTATTTTGAGG
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----TTTTGAGG
SEQ1616	-----
SEQ1617	-----
SEQ1601	TGATGATTGGAACATGCGATTAAACTCTCTTTAGAACTTTAAGTTTCCCAATTTTGG
SEQ1602	-----
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	TGATGATTGGAACATGCGATTAAACTCTCTTTAGAACTTTAAGTTTCCCAATTTTGG
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----ATGCGATTAAACTCTCTTTAGAACCTTTAAGTTTCCCAATTTTGG
SEQ1614	-----
SEQ1615	TGATGATTGGAACATGCGATTAAACTCTCTTTAGAACTTTAAGTTTCCCAATTTTGG
SEQ1616	-----
SEQ1617	-----
SEQ1601	AAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1602	-----AAACCGGGC
SEQ1603	-----TCGCTCTGCGGAACGTGAAGTATCTGTACTG-TCTGCAGAAA-GCGT
SEQ1604	-----
SEQ1605	-----AACGTGAAGTATCTGTACTGCTCTGCAGAAAAGCGT
SEQ1606	-----CTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1607	-----
SEQ1608	-----ATCTGTACTG-TCTGCAGAAAAGCGT
SEQ1609	AAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1610	-----TCTGTACTG-TCTGCAGAAA-GCGT
SEQ1611	AAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1612	-----CGTGAAGTATCTGTACTG-TCTGCAGAAA-GCGT
SEQ1613	AAACCCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1614	-----GT
SEQ1615	AAAACCGGCTAATATGGGGTCATCAGTAGGTATTTCAAAGCGACAGATGAATCCTCAC
SEQ1616	-----TGGTTCGCTCTGCGGAACGTGAAGTATCTGTACTG-TCTGCAGAAA-GCGT
SEQ1617	-----AAGCAGGGGATACATTGACCACTGAGTAAAACCGGGC
SEQ1601	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1602	TTGT-ATTCAGTTCGTTTAAAGAAGACTTGTCCATCTTTTCGTCAAAAAGAAATCACAGCG
SEQ1603	ATGC-GTGCTATTAATTATGATAAATTTTTGTTAAACTTATTTTATCACGCAAGTAG
SEQ1604	-----
SEQ1605	ATGC-GTGCTATTAATTATGATAAATTTTTGTTAAACTTATTTTATCACGCAAGTAG
SEQ1606	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1607	-----TTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1608	ATGC-GTGCTATTAATTATGATAAATTTTTGTTAAACTTATTTTATCACGCAAGTAG
SEQ1609	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG

Table 16: Comparative Sequences relating to SAG0767 (D-alanine – D-alanine ligase)

SEQ1610	ATGC-GTGCTATTAATTATGATAAATTTTTTGTAAAACCTATTTTATCACGCAAGTAG
SEQ1611	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1612	ATGC-GTGCTATTAATTATGATAAATTTTTTGTAAAACCTATTTTATCACGCAAGTAG
SEQ1613	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1614	ATGCCGTGCTATTAATTATGATAAATTTTTTGTAAAACCTATTTTATCACGCAAGTAG
SEQ1615	TCGCTCTGCAATTGACTTAGCTCTCAAGTATGATAGCCGTATTTTGATTGAACAAGGCG
SEQ1616	ATGC-GTGCTATTAATTATGATAAATTTTTTGTAAAACCTATTTTATCACGCAAGTAG
SEQ1617	TTGT-ATTCAGTTCGTTTAAAGAAGATCTGTCCATCTTTCGTCAAAAAGAAATCACAGCG
SEQ1601	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1602	GATAAACCACAAGC----CCCGATTGCTTTAAAAGCTTTACTTGCAATTTGACGCATTG
SEQ1603	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1604	-----
SEQ1605	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAA---
SEQ1606	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1607	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1608	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1609	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1610	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1611	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1612	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1613	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1614	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1615	GACAGCTCGTGAAATTGAAGTAGGTATTTTAGGCAATAATGATGTTAAGACAACCTTTTC
SEQ1616	GTCAATTTATTAAA---ACACAAGAATTTGATGAAATGCCATCTTCAGATGAAAAGTTA
SEQ1617	GATAAACCACAAGC----CCCGATTGCTTTAAAAGCTTTACTTGCAATTTGACGCATTG
SEQ1601	TGGCGAAGTTGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1602	TTCCATAGTT----GCTTCATCAACTTTAGCTGGAATATCCATAGTAATTTTATTATCA
SEQ1603	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1604	-----CGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1605	-----
SEQ1606	TGGCGAAGTCGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1607	TGGCGAAGTCGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1608	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1609	TGGCGAAGTCGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1610	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1611	TGGCGAAGTCGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1612	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1613	TGGCGAAGTTGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1614	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1615	TGGCGAAGTCGTTAAAGACGTCGATTTCTATGA--CTATGACGCCAAAT-ATATTGATA
SEQ1616	TGACAAACC-----AAACTGTTGATTTAGACAAAATGGTTCGTCCAAGTGATATCTATG
SEQ1617	TTCCATAGAT----GCTTCATCAACTTTAGCTGGAATATCCATAGCAATTTTATTATCA
SEQ1601	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1602	TATATTTGGCGTCATAGTCATAGAAATCGACGCTTTTAAAGACTTCGCCAGG--AAAAG
SEQ1603	TGATAATGCAAT--TGTTTTCCCGGTTTTAC---ATGGACCAATGGGGGAAG--ATGGT
SEQ1604	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1605	-----
SEQ1606	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1607	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1608	TGATAATGCAAT--TGTTTTCCCGGTTTTAC---ATGGACCAATGGGGGAAG--ATGGT
SEQ1609	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1610	TGATAAT
SEQ1611	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1612	TGATAATGCAAT--TGTTTTCCCGGTTTTAC---ATGGACCAATGGGGGAAG--ATGGT
SEQ1613	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1614	TGATAATGCAAT--TGTTTTCCCGGTTTTAC---ATGGACCAATGGGGGAAG--ATGGT
SEQ1615	TAAAATTACTAT--GGATATTCCAGCTAAAGTTGATGAAGCAACTATGGAAGCAATGCG
SEQ1616	TGATAATGCAAT--TGTTTTCCCGGTTTTAC---ATGGACCAATGGGGGAAG--ATGGT
SEQ1617	TATATTTGGCGT----ABLECMPARATIVESEQUENCESRELA-TINGTSAGD--ALANI

~~CONFIDENTIAL~~

Table 16: Comparative Sequences relating to SAG0767 (D-alanine - D-alanine ligase)

SEQ1601	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1602	T--GTCTTAACATCATTATTGCCTAAAATACCTACTTCAATTTACAGAGCTGTCAGCC
SEQ1603	C--TATCCAAG--GATTTTTAGAAAGTTTTAAGGATGCCTTATGTTGGGACTAATATTCT
SEQ1604	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1605	-----
SEQ1606	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1607	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1608	C--TATCCAAG--GATTTTTAGAAAGTTTTAAGGATGCCTTATGTTGGGACTAATATTCT
SEQ1609	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1610	-----
SEQ1611	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1612	C--TATCCAAG--GATTTTTAGAAAGTTTTAAGGATGCCTTATGTTGGGACTAATATTCT
SEQ1613	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1614	C--TATCCAAG--GATTTTTAGAAAGTTTTAAGGATGCCTTATGTTGGGACTAATATTCT
SEQ1615	CAATATGCAAGTAAAGCTTTTAAAGCAATCGGGGCTTGTGGTTTATCACGCTGTGATTT
SEQ1616	C--TATCCAAG--GATTTTTAGAAAGTTTTAAGGATGCCTTATGTTGGGACTAATATTCT
SEQ1617	E--DALANINELIGASE-----
SEQ1601	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1602	TGTTCAATCAAAATACGGCTATCATACTTGAGAGCTAAGTCAATKSCAGAGCGAAGTGA
SEQ1603	TCTTCAAGCGTGGCTAT-----
SEQ1604	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1605	-----
SEQ1606	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1607	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1608	TCTTCAA-----
SEQ1609	TTTTTGACGAAAGAA-TGGACAAATCTTCTTAAACGAACTGAAATAC-----
SEQ1610	-----
SEQ1611	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1612	TCTTCAAGCGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAG----
SEQ1613	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1614	TCTTCAAGCGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAGGTGT
SEQ1615	TTTTTGACGAAAGA--TGGACAAATCTTCTTAAACGAACTGAA-TACAATGCCCGGTTT
SEQ1616	TCTTCAAGCGTGGCTATGGATAAAATTACAACAAAACAAGTCCTTGCAACAGTAGG----
SEQ1617	-----
SEQ1601	ACTCAGTGGTCAATGTATCCTCTGCTTTGGGAAAA-TATGGGGCTAACTTATAGTGATT
SEQ1602	GATTCATCTGTCGCTTTTGAAATACCTACTGATGACCCCATATTAGCCGGTTTACAAA
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	ACTCAGTGGTCAATGTATCCTCTGCTTTGGGAAAA-T-----
SEQ1607	ACTCAGTGGTCAATGTATCCCCTGCTTTGGGAAAAGTATGGGGCTAACCTT-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	ACTCAGTGGTCAATGTATCCCCTGCTTTGGGAAAA-TATGGGGCTAACTTATAG-----
SEQ1612	-----
SEQ1613	ACTCAGTGGTCAATGTATCCTCTGCTTTGGGAAAA-TATGGGGCTAACTT-----
SEQ1614	CCTCAGG-----
SEQ1615	ACTCAGTGGTCAATGTATCCCCTGCTTTGGGAAAA-TATGGGGCTAACTTATAGTGA--
SEQ1616	-----
SEQ1617	-----

Table 16: Comparative Sequences relating to SAG0767 (D-alanine - D-alanine ligase)

SEQ1601	GATTG-----
SEQ1602	ATTGGGAAACTTAAAGTTTCTAAAGAGAGTTTAATCGCATGTTCCAAATCATCACCCCTC
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	-----
SEQ1602	AAATAAGTTTGATATGCAACCTGAGGTACACCTACTGTTGCAAGGACTTGTTTTGTGTG
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	-----
SEQ1602	ATTTTATCCATAGCCACGCTTGAAGATAGAATATTAGTCCCAACATAAGGCATCCTTAA
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

Table 16: Comparativ Sequences relating t SAG0767 (D-alanine - D-alanine ligase)

SEQ1601	-----
SEQ1602	ACTTCTAAAAATCCTTGGATAGAACCATCTTCCCCCATTGGTCCATGTAAAACGGGGAA
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	-----
SEQ1602	ACAATTGCATTATCATCATAGATATCACTTGGACGAACCATTTTGTCTAAATCAACAGT
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----
SEQ1601	-----
SEQ1602	TGGTTTGTCAATTAACCTTTTCATCTGAAGATGGCATTTCATCAAATTCTTGTGTTTTAAT
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

Table 16: Comparative Sequences relating to SAG0767 (D-alanine – D-alanine ligase)

SEQ1601	-----
SEQ1602	AATTGACCTACTTGCGTG
SEQ1603	-----
SEQ1604	-----
SEQ1605	-----
SEQ1606	-----
SEQ1607	-----
SEQ1608	-----
SEQ1609	-----
SEQ1610	-----
SEQ1611	-----
SEQ1612	-----
SEQ1613	-----
SEQ1614	-----
SEQ1615	-----
SEQ1616	-----
SEQ1617	-----

Table 17: Comparative Sequences relating to SAG1086 (xanthine ph ph rib syltransferase)

SEQ ID NO. 1701: SAG1086 FROM THE1169NT1 GBS NONTYPEABLE STRAIN

TTTAAAGGTTGATTCTTTTGGACTCATCAGGTAGATTTTGGAGTTAATGCAGGAAATAGGTAAAGTTTTTGCTGATAAATATAAAGAAG
CCGGCATTACGAAGGTTGTTACGATTGAAGCATCTGGAATTGCGCCAGCAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTT
GCTAAAAAGGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGCAAGWTACGAGTCAAGTTTC
TATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACGGTCAAGCGGCTAAAGGATTACTTG
AAATTATTGGTCAAGCTGGAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATGGGCGTGATTGTAGAAAAA
ACAGGTGTTCCAGT

SEQ ID NO. 1702: SAG0767 FROM THE 18RS21 GBS TYPE II STRAIN

TTTAGGTGAGAACATTTTAAAGGTTGATTCTTTTGGACTCATCAGGTAGATTTTGGAGTTAATGCAGGAAATAGGTAAAGTTTTTGCTG
ATAAATATAAAGAAGCCGGCATTACGAAGGTTGTTACGATTGAAGCATCTGGAATTGCACCAGCAGTGTACGCAGCTCAAGCATTGGGC
GKACCAATGATATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGCAAGT
TACGAGTCAAGTTCTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACGGTCAAGCGG
CTAAAGGATTACTTGAAATTATTGGTCAAGCTGGAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATGGGCGT
GATTGTAGAAAAACA

SEQ ID NO. 1703: SAG0767 FROM THE H36b1 GBS TYPE Ib STRAIN

AAGAAGCTATTCTTAAAGATGGTGTATGTTTTAGGTGAGAACATTTTAAAGTTGATTCTTTTGGACTCATCAGGTAGATTTTGGAGTTA
ATGCAGGAAATAGGTAAAGTTTTTGCTGATAAATATAAAGAAGCCGGCATTACGAAGGTTGTTACAATTGAAGCATCTGGAATTGCGCC
AGCAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGT
CTGAAGTGTATTTCTTTTACAAAGCAAGTTACGAGTCAAGTTTCTATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCAT
GATGACTTTTACAAACGGTCAAGCGGCTAAGGATTACTTGAAATTATTGGTCAAGCTGGAGCTAAGGTTGCTGGTATCGGAATCYT
TATTGAAAAATCTTCCAGATGGGCGTGATT

SEQ ID NO. 1704: SAG0767 FROM THE M732 GBS TYPE III STRAIN

ATTCTTTTGGACTATCAGGTAAATTTTGGAGTTAATGCAGGAAATAGGTAAAGTTTTTGCTGATAAATATAAAGAAGCCGGCATTACGA
AGGTTGTTACAATTGAAGCATCTGGAATTGCGCCAGCAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTTGCTAAAAAGCT
AAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGCAAGTTACGAGTCAAGTTTCTATTGTGAGTCG
CTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACGGTCAAGCGGCTAAGGATTACTTGAAATTATTGGTC
AAGCTGAAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATGGGCGTGATTGTAGAAAAACAGGTGTTCCG
GTTACTTCTCTTGCTCGT

SEQ ID NO. 1705: SAG0767 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GAACGTATTCTTAAAGATGGTGTATGTTTTAGGTGAGAACATTTTAAAGTTGATTCTTTTGGACTCATCAGGTAAATTTGAGTTAAT
GCAGGAAATAGGTAAAGTTTTGCTGATAAATATAAAGAAGCCGGCATTACGAAGGTTGTTACAATTGAAGCATCTGGAATTGCGCCAG
CAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATRTTAAGTGT
GAAGTGTATTCTTTTACAAAGCAAGTTACGAGTCAAGTTTCTATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGA
TGACTTTTAAACAAACGGTCAAGC

SEQ ID NO. 1706: SAG0767 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

ACATTTTAAAGGTTGATTCTTTTGGACTCATCAGGTAGATTTTGGAGTTAATGCAGGAAATAGGTAAAGTTTTTGCTGATAAATATAAA
GAAGCCGGCATTACGAAGGTTGTTACGATTGAAGCATCTGGAATTGCACCAGCAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGAT
ATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGCAAGTTACGAGTCAAG
TTTCTATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACMGTCYAGCGGCTAAAGGATTA
CTTGAAATTATTGGTCAAGCTGGAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATGGGCGTGATTGTAGAAA
AAA

SEQ ID NO. 1707: SAG0767 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

ACGTATTCTTAAAGATGGTGTATGTTTTAGGTGAGAACATTTTAAAGTTGATTCTTTTGGACTCATCAGGTAGATTTTGGAGTTAATGC
AGGAAATAGGTAAAGTTTTGCTGATAAATATAAAGAAGCCGGCATTACGAAGGTTGTTACAATTGAAGCATCTGGAATTGCGCCAGCA
GTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGT
AGTGTATTCTTTTACAAAGCAAGTTACGAGTCAAGTTTCTATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGATG
ACTTTTACAAACGGKCAAGCGGSTAAAGGATTACTTGAAATTATTGGTCAAGCTGGAGCTA

SEQ ID NO. 1708: SAG0767 FROM THE COH1 GBS TYPE Ia STRAIN

TTTAAAGTTGATTCTTTTGGACTCATCAGGTAAATTTTGGAGTTAATGCAGGAAATAGGTAAAGTTTTTGCTGATAAATATAAAGAAG
CCGGCATTACGAAGGTTGTTACAATTGAAGCATCTGGAATTGCGCCAGCAGTGTACGCAGCTCAAGCATTGGGCGTACCAATGATATTT
GCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGCAAGTTACGAGTCAAGTTTC
TATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACGGTCAAGCGGCTAAAGGATTACTTG
AAATTATTGGTCAAGCTGAAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATGGGCGTGATTGTAGAAAAA
ACAGGTGTTCCGGTTAC

SEQ ID NO. 1709: SAG0767 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

GCTGATAAATATAAAGAAGCCGGCATTACGAAGGTTGTTACAATTGAAGCATCTGGAATTGCGCCAGCAGTGTACGCAGCTCAAGCATT
GGGCGTACCAATGATATTTGCTAAAAAGCTAAGAACATTACTATGACTGAAGGTATCTTAAGTGTGAGTGTATTCTTTTACAAAGC
AAGTTACGAGTCAAGTTTCTATTGTGAGTCGCTTTTATCTAACGATGATACTGTACTCATCATTTGATGACTTTTACAAACGGTCAA
GCGGCTAAAGGATTACTTGAAATTTATTGGTCAAGCTGGAGCTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTCCAGATG
GGCGTGATTGTAGAAAAACAGGTGTTCCAGT

Table 17: Comparative Sequences relating t SAG1086 (xanthine phosphoribosyltransferase)

SEQ1701	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGWTACGA
SEQ1702	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1703	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1704	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1705	TTACTATGACTGAAGGTATRTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1706	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1707	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1708	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1709	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1710	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1711	TTACTATGACTGAAGGTATCTTAACTGCTGAAGTGTATTCTTTTACAAAGCAAGTTACGA
SEQ1701	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1702	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1703	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1704	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1705	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1706	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1707	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1708	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1709	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1710	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1711	GTCAAGTTTCTATTGTGAGTCGCTTTTTATCTAACGATGATACTGTACTCATCATTGATG
SEQ1701	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1702	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1703	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1704	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1705	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1706	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1707	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1708	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1709	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1710	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1711	ACTTTTGTAGCAAACGGTCAAGCGGCTAAAGGATTACTTGAAATT-ATTGGTCAAGCTGGA
SEQ1701	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1702	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1703	CTAAGGTTGCTGGTATCGGAATCYTTATTGAAAAATCTTTCCAAGATGGGCGTGATT--
SEQ1704	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1705	-----
SEQ1706	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1707	CTA-----
SEQ1708	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1709	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1710	CTAAGGTTGCTGGTATCGGAATCGTTATTGAAAAATCTTTCCAAGATGGGCGTGATTG
SEQ1711	CTAAGGTTGCTGGTATCGGA-----TABCMARATVSTNCSR--ATNGTSAGXANTHN
SEQ1701	TAGAAAAAACAGGTGTTCCAGT-----
SEQ1702	TAGAAAAACA-----
SEQ1703	-----
SEQ1704	TAGAAAAAACAGGTGTTCCGGTTACTTCTCTTGCTCGT
SEQ1705	-----
SEQ1706	TAGAAAA-----
SEQ1707	-----
SEQ1708	TAGAAAAAACAGGTGTTCCGGTTAC-----
SEQ1709	TAGAAAAAACAGGTGTTCCAGT-----
SEQ1710	TAGAAAAAACAGGTGTTCCAG-----
SEQ1711	HRBSYTRANSRAS-----

Table 18: Comparativ Sequences relating t SAG1600 (glutamate racemase)

SEQ ID NO. 1801: SAG1600 FROM THE H36b GBS TYPE Ib STRAIN (REVERSE COMPLEMENT)
 AATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTTCT
 TATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTGCCTGGCAAGAAATTAAGAAAAACTAGACGTG
 CCTGTTTTAGGCGTTATTTTACCAGGAGCTAGCGCAGCTATCAAATCAACTAATTCAGGGAAAGTTGGTATTATAGGTACTCCCATGAC
 TGTTAAATCAGATGCTTATCGTCAAAAAATTCAGCTTTGTCTCAAATACTGCTGTGGTATCCCTTGCTTGTCCGAAATTTGTTCCAA
 TTGTGGAATCAAATCAGATGTCTTCTAGTTTAGCCAAAAAGGTGGTTTATGAAACGTTGTCCCATTAGTTGGTAAATTAGATACTTTA
 ATTTTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGGGCTGAGGTTAAATTAATTGATAGTGGCGCAGA
 AACCGTTCGTGATATTTCTGTTTTATTGAACATTTTGGAGATAAACCATTAATTGGCAAAATAAACACGGTGGTCATCACTTTTACACAA
 CCGCCAGCCCAA

SEQ ID NO. 1802: SAG1600 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
 AAATGTTCCGTCAACTTCCAGAAGAGGAAGTAATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATT
 AGAGAGTTTACCTGGCAGATGGTTAACTTCTTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTGC
 CTGGCAAGAAATTAAGAAAAACTAGACATCCCTGTTTTAGGCGTTTATGCTGTTAAATCAGATGCTTATCGTCAAAAAATTCAGCTTTGTCTCAAATACTGCTGTG
 GGAAGTTGGTATTATAGGTACTCCCATGACTGTTAAATCAGATGCTTATCGTCAAAAAATTCAGCTTTGTCTCAAATACTGCTGTG
 GTATCCCTTGCTTGTCCGAAATTTGTTCCAATTGTGGAATCAAATCAGATGTCTTCTAGTTTAGCCAAAAAGGTGGTTTATGAAACGTT
 GTCCCATTAGTTGGTAAATTAGATACTTTAATTTTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGG
 CTGAGGTTAAATTAATTGATAGTGGCGCAGAAACCGTTCGTGATATTTCTGTTTTATTGAACATTTTGGAGATAAACCATTAATTGGCAA
 AATAAACACGGTGGTCATCACTTTTACACAACCGCCAGCCCAAAAGGTTTAAAGAAA

SEQ ID NO. 1803: SAG1600 FROM THE 090 GBS TYPE Ia STRAIN
 AATCTTCATTGGAGACCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATTAGAGAGTTACCTGGCAGATGGTTAATTTCTT
 ATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTGCCTGGCAAGAAATTAAGAAAAACTAGACATAC
 CTGTTTTAGGCGTTATTTTACCAGGAGCTAGCGCAGCTATCAAATCAACTAATTCAGGGAAAGTTGGTATTATAGGTACTCCCATGACT
 GTTAAATCAGATGCTTATCGTCAAAAAATTCAGCTTTGTCTCAAATACTGCTGTGGTATCCCTTGCTTGTCCGAAATTTGTTCCAAT
 TGTGGAATCAAATCAGATGTCTTCTAGTTTAGCCAAAAAGGTGGTTTATGAAACGCTGTCCCATTAGTTGGTAAATTAGATACTTTAA
 TTTTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGGGCTGAGGTTAAATTAATTGATAGTGGCGCAGAA
 ACCGTTTCGTGATATTTCTGTTTTATTGAACATTTTGGAGATAmCCATaATTGGsmaAATAAACACGGTGGTCATCACTTTTACACAAC
 CGsCAGCCCAAAAGGTTTTTAAGGAAATTGCAGAACATGGCTTAATCAAGAAATAAAT

SEQ ID NO. 1804: SAG1600 FROM THE A909 GBS TYPE Ia STRAIN
 GCGGTTGTGTAAAGTGATGACCACCGTGTATTTTGGCAATTATGGTTTATCTCAAAATAGTTCAATAAAACAGAAATATCACGAAC
 GGTTCCTGCGCCACTATCAATTAATTTAACCTCAGCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGCGTGCAACCTA
 AAATTAAGTATCTAATTTACCAACTAATGGGGACACGTTTCATAAACACCTTTTGGCTAAACTAGAAGACATCTGATTGATTCC
 ACAATTGGAACAAATTCGGAACAAGCAAGGGATACCAACAGTATTTGGAGACAAAGCTTGAATTTTTTACGATAAGCATCTGATTT
 AACAGTCATGGGAGTACCTATAATACCAACTTTCCCTAAATTAGTTGATTGATAGCTGCGCTAGCTCCTGGTAAATAACGCCATAAA
 CAGGGATGTCTAGTTTTCTTAAATTTCTTGCCAGGCAACTGCAGTTGCTGTATTACAAGCTATAACAATCATCTTAACATTTTATGTC
 AATAAGAAGTTAACCATCTGCCAGGTAACTCTCTAATCTGTTGAGCAGGCTAGGACCATACGGAGCTCTAGCCTGATCTCCAATGAA
 GATTACTTCTCTTCTGGAAGTTGACGGAACATTTCCCTTAACAACCGTTAAACACCT

SEQ ID NO. 1805: SAG1600 FROM THE COH1 GBS TYPE Ia STRAIN
 TTCCGTCAACTTCCAAATATGAAGTAATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATTAGAGA
 GTTTACCTGGCAGATGGTTAACTTCTTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTGCCTGGC
 AAGAAATTAAGAAAAACTAGACATCCCTGTTTTAGGCGTTATTTTACCAGGAGCTAGCGCAGCTATCAAATCAACTAATTTAGGGAAA
 GTTGGTATTATAGGTACTCCCATGACTGTTAAATCAGATGCTTATCGTCAAAAAATTCAGCTTTGTCTCAAATACTGCTGTGGTATC
 CCTGCTTGTCCGAAAT

SEQ ID NO. 1806: SAG1600 FROM THE CJB110 GBS NONTYPEABLE STRAIN
 GTAATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
 CTTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTGCCTGGCAAGAAATTAAGAAAAACTAGACA
 TAC

SEQ ID NO. 1807: SAG1600 FROM THE 1169NT1 GBS TYPE V STRAIN
 CTTTTGGGCTGGCGGTTGTGTAAATTTGATGACCACCGTGTATTTTGGCAATTATGGTTTATCTCAAAATAGTTCAATAAAACAGAA
 ATATCAGCAACGGTTTCTGCGCCACTATCAATTAATTTAACCTCAGCCCCATAACATTTTGAATAATGGGACGTAATAGGGGATAATG
 CGTGCAACCTAAATTAAGTATCTAATTTACCAACTAATGGGGACAATGTTTCATAAACACCTTTTGGCTAAACTAGAAGACATCT
 GATTGTATCCCAATTTGGAACAAATTCGGAACAAGGGATACCACAGCAGTATTTGGAGACAAAGCTTGAATTTTTTGGACGATAA
 GCATCTGATTTAACAGTCATGGGAGTACCTATAA

SEQ ID NO. 1808: SAG1600 FROM THE 1169NT1 GBS TYPE V STRAIN
 GTAATCTTCATTGGGGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
 CTTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTT

SEQ ID NO. 1809: SAG1600 FROM THE 18RS21 GBS TYPE II STRAIN
 GAAATGTTCCGTCAACTTCCAGAAGAGGAAGTAATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCCTAGACCTGCTCAACAGAT
 TAGAGAGTTTACCTGGCAGATGGTTAACTTCTTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAATACAGCAACTGCAGTTG
 CCTGGCAAGAAATTAAGAAAAACTAGACATCCCTGTTTTAGGCGTTATTTTACCAGGAGCTAGCGCAGCTATCAAATCAACTAATTTA
 GGGAAAGTTGGTATTATAGGTACTCCCATGACTGTTAAATCAGATGCTTATCGTCAAAAAATTCAGC

Table 18: Comparative Sequences relating to SAG1600 (glutamate racemase)

SEQ ID NO. 1810: SAG1600 FROM THE 18RS21 TYPE II STRAIN

ATTTCTTTAAAACCTTTTGGGCTGGCGGTTGTGTAATAATGATGACCACCGTGTATTTTGGCCAATTATGGTTTATCTCAAAATAGT
CAATAAAACAGAAATATCACGAACGGTTTCTGCGCCACTATCAATTAATTTAACCTCAGCCCCCATAACATTTGAATGATGGGACGTA
ATATGGGATAATGCGTGCAACCTAAAATTAAAGTA

SEQ ID NO. 1811: SAG1600 FROM THE 2603 V/R GBS TYPE V STRAIN

ATTTCTTTAAAACCTTTTGGGCTGGCGGTTGTGTAATAAGTATGACCACCGTGTATTTTGGCCAATTATGGTTTATCTCAAAATAGT
TCAATAAAACAGAAATATCACGAACGGTTTCTGCGCCACTATCAATTAATTTAACCTCAGCCCCCATAACATTTGAATGATGGGACGT
AATAGGGGATAATGCGTGCAACCTAAAATTAAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTGGCTAA
ACTAGAAGACATCTGATTTGATTCACAATTGGAACA

SEQ ID NO. 1812: SAG1600 FROM THE M781 GBS TYPE III STRAIN

GGCGGTTGTGTAAGTATGACCACCGTGTATTTTGGCCAATTATGGTTTATCTCAAAATAGTTCAATAAAACAGAAATATCACGAA
CGGTTTCTGCGCCACTATCAATTAATTTAACCTCAGCCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGCGTGCAACCT
AAAATTAAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTGGCTAACTAGAAGA

SEQ ID NO. 1813: SAG1600 FROM THE M 781 GBS TYPE III STRAIN

AATCTTCATTGGAGATCAGGCTAGAGCTCCGTATGGTCTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAACTTCT
TATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGAATACAGCAACTGC

SEQ ID NO. 1814: SAG1600 FROM THE JM9130013 GS TYPE VIII STRAIN

TGGGCTGGCGGTTGTGTAAGTATGACCACCGTGTATTTTGGCCAATTATGGTTTATCTCAAAATAGTTCAATAAAACAGAAATAT
CACGAACGGTTTCTGCGCCACTATCAATTAATTTAACCTCAGCCCCCATAACATTTTGAATGATGGGACGTAATAAGGGATAATGCGTG
CAACCTAAAATTAAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTGGCTAACTAGAAGACATCTGATT
TGATTCCACAATTGGAACAAATTTGCGACAAGCAAGGGATACCACAGCAGTATTTGGAGACAAAGCTTGAATTTTGTACGATAAGCAT
CTGATTTAACAGTCATGGGAGTACCTATAATACCAACTTTCCCTGAA

SEQ1801	-----AATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1802	AAATGTTCCGTCAACTTCCAGAAGAGGAAGTAATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1803	-----AATCTTCATTGGAGACCAGGCTAGAGCT
SEQ1804	-----GCGGTTGTGTAAG-T
SEQ1805	-----TTCCGTCAACTTCCAAAATATGAAGTAATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1806	-----GTAATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1807	-----CTTTTGGGCTGGCGGTTGTGTAAG-T
SEQ1808	-----GTAATCTTCATTGGGGATCAGGCTAGAGCT
SEQ1809	AAATGTTCCGTCAACTTCCAGAAGAGGAAGTAATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1810	-----ATTTCTTTAAAACCTTTTGGGCTGGCGGTTGTGTAATAT-T
SEQ1811	-----ATTTCTTTAAAACCTTTTGGGCTGGCGGTTGTGTAATAAGT
SEQ1812	-----GGCGGTTGTGTAAG-T
SEQ1813	-----AATCTTCATTGGAGATCAGGCTAGAGCT
SEQ1814	-----TGGGCTGGCGGTTGTGTAAG-T

SEQ1801	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1802	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1803	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTT-ACCTGGCAGATGGTTAATTT
SEQ1804	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---
SEQ1805	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1806	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1807	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---
SEQ1808	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1809	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1810	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---
SEQ1811	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---
SEQ1812	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---
SEQ1813	CGTATGGTC-CTAGACCTGCTCAACAGATTAGAGAGTTTACCTGGCAGATGGTTAATTT
SEQ1814	--GATGACCACCGTGTTTATTTTGGCCAATTATGG--TTTATCTCA-AAATAGTTCA---

Table 18: Comparative S equences relating to SAG1600 (glutamate racemase)

SEQ1801	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1802	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1803	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1804	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1805	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1806	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1807	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1808	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTT--
SEQ1809	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGCAGTTGC
SEQ1810	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1811	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1812	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1813	TTATTGACTAAAAATGTTAAGATGATTGTTATAGCTTGTAAATACAGCAACTGC-----
SEQ1814	--ATAAACAGAAATATCACGAACGGT-TTCTGCGCCACTATCAATTAATTTAACCTCA
SEQ1801	TGGCAAGAAATTAAAGAAAAACTAGACGTGCCTGTTTTAGGCGTTATTTTACCAGGAGC
SEQ1802	TGGCAAGAAATTAAAGAAAAACTAGACATCCCTGTTTTAGGCGTTATTTTACCAGGAGC
SEQ1803	TGGCAAGAAATTAAAGAAAAACTAGACATACCTGTTTTAGGCGTTATTTTACCAGGAGC
SEQ1804	CCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGC-GTGCAACCTAAAAAT
SEQ1805	TGGCAAGAAATTAAAGAAAAACTAGACATCCCTGTTTTAGGCGTTATTTTACCAGGAGC
SEQ1806	TGGCAAGAAATTAAAGAAAAACTAGACATAC-----
SEQ1807	CCCCCATAACATTTTGAATAATGGGACGTAATAGGGGATAATGC-GTGCAACCTAAAAAT
SEQ1808	-----
SEQ1809	TGGCAAGAAATTAAAGAAAAACTAGACATCCCTGTTTTAGGCGTTATTTTACCAGGAGC
SEQ1810	CCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGC-GTGCAACCTAAAAAT
SEQ1811	CCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGC-GTGCAACCTAAAAAT
SEQ1812	CCCCCATAACATTTTGAATGATGGGACGTAATAGGGGATAATGC-GTGCAACCTAAAAAT
SEQ1813	-----
SEQ1814	CCCCCATAACATTTTGAATGATGGGACGTAATAAGGGATAATGC-GTGCAACCTAAAAAT
SEQ1801	AGCGCAGCTATCAAATCAACTAATTAGGGGAAAGTTGGTATTATAGGTACTCCCATGAC
SEQ1802	AGCGCAGCTATCAAATCAACTAATTAGGGGAAAGTTGGTATTATAGGTACTCCCATGAC
SEQ1803	AGCGCAGCTATCAAATCAACTAATTAGGGGAAAGTTGGTATTATAGGTACTCCCATGAC
SEQ1804	AAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTTGGCTAA
SEQ1805	AGCGCAGCTATCAAATCAACTAATTAGGGGAAAGTTGGTATTATAGGTACTCCCATGAC
SEQ1806	-----
SEQ1807	AAAGTATCTAATTTACCAACTAATGGGGACAATGTTTCATAAACCACCTTTTTGGCTAA
SEQ1808	-----
SEQ1809	AGCGCAGCTATCAAATCAACTAATTAGGGGAAAGTTGGTATTATAGGTACTCCCATGAC
SEQ1810	AAAGTA-----
SEQ1811	AAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTTGGCTAA
SEQ1812	AAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTTGGCTAA
SEQ1813	-----
SEQ1814	AAAGTATCTAATTTACCAACTAATGGGGACAACGTTTCATAAACCACCTTTTTGGCTAA
SEQ1801	GTTAAATCAGATGCTTATCGTCAAAAAATTCAAGCTTTGTCTCCAAATACTGCTGTGGT
SEQ1802	GTTAAATCAGATGCTTATCGTCAAAAAATTCAAGCTTTGTCTCCAAATACTGCTGTGGT
SEQ1803	GTTAAATCAGATGCTTATCGTCAAAAAATTCAAGCTTTGTCTCCAAATACTGCTGTGGT
SEQ1804	CTAGAAGACATCTGATTTGATTCCACAATTGGAACAAATTTCCGACAAGCAAGGGATAC
SEQ1805	GTTAAATCAGATGCTTATCGTCAAAAAATTCAAGCTTTGTCTCCAAATACTGCTGTGGT
SEQ1806	-----
SEQ1807	CTAGAAGACATCTGATTTGATTCCACAATTGGAACAAATTTCCGACAAGCAAGGGATAC
SEQ1808	-----
SEQ1809	GTTAAATCAGATGCTTATCGTCAAAAAATTCAAGC-----
SEQ1810	-----
SEQ1811	CTAGAAGACATCTGATTTGATTCCACAATTGGAACAA-----
SEQ1812	CTAGAAGA-----
SEQ1813	-----
SEQ1814	CTAGAAGACATCTGATTTGATTCCACAATTGGAACAAATTTCCGACAAGCAAGGGATAC

Table 18: Comparative Sequences relating to SAG1600 (glutamate racemase)

SEQ1801	TCCCTTGCTTGTCCGAAATTTGTTCCAATTGTGGAATCAAATCAGATGTCTTCTAGTTT
SEQ1802	TCCCTTGCTTGTCCGAAATTTGTTCCAATTGTGGAATCAAATCAGATGTCTTCTAGTTT
SEQ1803	TCCCTTGCTTGTCCGAAATTTGTTCCAATTGTGGAATCAAATCAGATGTCTTCTAGTTT
SEQ1804	ACAGCAGTATTTGGAGACAAAGCTTGAATTTTTTGACGATAAGCATCTGATTAAACAGT
SEQ1805	TCCCTTGCTTGTCCGAAAT-----
SEQ1806	-----
SEQ1807	ACAGCAGTATTTGGAGACAAAGCTTGAATTTTTTGACGATAAGCATCTGATTAAACAGT
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	ACAGCAGTATTTGGAGACAAAGCTTGAATTTTTTGACGATAAGCATCTGATTAAACAGT
SEQ1801	GCCAAAAAGGTGGTTTATGAAACGTTGTCCCCATTAGTTGGTAAATTAGATACTTTAAT
SEQ1802	GCCAAAAAGGTGGTTTATGAAACGTTGTCCCCATTAGTTGGTAAATTAGATACTTTAAT
SEQ1803	GCCAAAAAGGTGGTTTATGAAACGTTGTCCCCATTAGTTGGTAAATTAGATACTTTAAT
SEQ1804	ATGGGAGTACCTATAATACCAACTTTCCCTAAATTAGTTGATTGATAGCTGCGCTAGC
SEQ1805	-----
SEQ1806	-----
SEQ1807	ATGGGAGTACCTATAA-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	ATGGGAGTACCTATAATACCAACTTTCCCTGAATABCMARATVSTNCSRATNGTSAGGT
SEQ1801	TTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGGGCTGA
SEQ1802	TTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGGGCTGA
SEQ1803	TTAGGTTGCACGCATTATCCCTTATTACGTCCCATCATTCAAATGTTATGGGGGCTGA
SEQ1804	CCTGGTAAATAACGCCTAAACAGGGATGTCTAGTTTTTCTTTAATTTCTTGCCAGGC
SEQ1805	-----
SEQ1806	-----
SEQ1807	-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	AMATRACMAS-----
SEQ1801	GTTAAATTAATTGATAGTGGCGCAGAAACCGTTCGTGATATTTCTGTTTTATTGAACTA
SEQ1802	GTTAAATTAATTGATAGTGGCGCAGAAACCGTTCGTGATATTTCTGTTTTATTGAACTA
SEQ1803	GTTAAATTAATTGATAGTGGCGCAGAAACCGTTCGTGATATTTCTGTTTTATTGAACTA
SEQ1804	ACTGCAGTTGCTGTATTACAAGCTATAACAATCATCTTAACATTTTAGTCAATAAGAA
SEQ1805	-----
SEQ1806	-----
SEQ1807	-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	-----

Table 18: Comparative Sequences relating to SAG1600 (glutamate racemase)

SEQ1801	TTTGAGATAAACCATAATTGGCAAAATAAACACGGTGGTCATCACTTTTACACAACCGC
SEQ1802	TTTGAGATAAACCATAATTGGCAAAATAAACACGGTGGTCATCACTTTTACACAACCGC
SEQ1803	TTTGAGATAAMCCATAATTGGSMAAATAAACACGGTGGTCATCACTTTTACACAACCGS
SEQ1804	TTAACCATCTGCCAGGTAAACTCTCTAATCTGTTGAGCAGGTCTAGGACCATACGGGAGC
SEQ1805	-----
SEQ1806	-----
SEQ1807	-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	-----
SEQ1801	AGCCCAA-----
SEQ1802	AGCCCAAAGGTTTTAAAGAAA-----
SEQ1803	AGCCCAAAGGTTTTTAAGGAAATTGCAGAACAATGGCTTAATCAAGAAATAAAT----
SEQ1804	CTAGCCTGATCTCCAATGAAGATTACTTCCTCTTCTGGAAGTTGACGGAACATTTTCCTT
SEQ1805	-----
SEQ1806	-----
SEQ1807	-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	-----
SEQ1801	-----
SEQ1802	-----
SEQ1803	-----
SEQ1804	ACAACCGTTAAACCACCT
SEQ1805	-----
SEQ1806	-----
SEQ1807	-----
SEQ1808	-----
SEQ1809	-----
SEQ1810	-----
SEQ1811	-----
SEQ1812	-----
SEQ1813	-----
SEQ1814	-----

Table 19: Comparative Sequences relating to SAG1680 (shikimate 5-dehydratase)

SEQ ID NO. 1901: SAG1680 FROM THE 2603 V/R GBS TYPE V STRAIN

ATCCCTAGACCATTATAAGCATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAAC
CAAGTCGACAACACTACTAAATTCGGTGTTAAAATTTCTGGATCGTTAATTAACTATAATTATCTAATGGCCTCATTCCCTA
AAGTAGTAGCATCAATATAAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACT
ACCTTTATTTGAACTGTTTTTAAATTTTATCTGATAAGTCAATGACCTTATCGTAATTTGAGCTGTTACGATTAAATAA
TCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAGTCAACTGCTTTACCTGAACCAATACCAGCTATTGTAATTA
TTTTATTTTTCAGCACTGAAACCTTGAGCTGCTAAAGCTTTAAAACAACCAATGCCATCTGTCATATGGCCTACTAAACGT
CCGGTTCCACCTTGATTAAACGATAGTATTTACAGCACCCTAATTTAGCTTGAGGAGATAAATCATCTAGCAAAGGGAT
AACACTCTGTTTAAATGGCATTGAAACATTAACACCACGAATACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATT
TACCCTCTTCTACTTCAAATGTCAGATAGGCATAATTCATGTTTTTTCTTGAAAAGAGGTATTCCACATTAAACGGGGAT
AGAGAGTGGCGTGCAGG

SEQ ID NO. 1902: SAG1680 FROM THE H36b GBS TYPE Ib STRAIN

GTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCTAGACCATTATAAGCATGTTTCACTCCATTTTGTCT
AACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCAAGTCGACAACACTACTAAATTCGGTGTTAAAATTTCTGGAT
CGTTAATTAACTATAATTATCTAATGGCCTCATTCTTAACTAGTAGCATCAATATAAAAAATGACTAGTTCTAATAGCG
TCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTACCTTTATTTGAACTGTTTTTAAATTTATCTGATAAGTC
AATGACCTTATCGTAATTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAGTGCAA
CTGCTTTACCTGAACCAATACCAGCTATTGTAATTTATTTTATTTTTCAGCACTGAAACCTTGAGCTGCTAAAGCTTTA
AAACAACCAATGCCATCTGTCATATGGCCTACTAAACGTCGGTTCCACCTTGATTACGATAGTATTACAGCACCAC
TAATTTAGCTTGAGGAGATAAATCATCTAGCAAAGGGATAACACTCTGTTTAAATGGCATTGAAACATTAACACCAGAA
TACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTACCCTCTTCTACTTCAAATGTCAGATAGGCATAATTCATG
TTTTTTCTTGAAAAGAGGTATTCCACATTAAACGGGGATAGAGAGTGGCGTGCAGGA

SEQ ID NO. 1903: SAG1680 FROM THE M732 GBS TYPE III STRAIN

CTGGTCTAATTGCCAATCCTGCACGCCACTCTCTATCCCCGTTAATGTGGAATACCTCTTTTCAAGAAAAAACATGAAT
TATGCCATCTGACATTTGAAGTAGAAGAGGGTAAATTAACAGAAGCTGTTTCGAGGTGTCAGGGCATTGAGTATTCGTGG
TGTTAATGTTTCAATGCCATTTAAACAGAGTGTTATCCCTTTGCTAGATGATTTATCTCCTCAAGCTAAATTAGTGGGTG
CTGTAAATACTATCGTTAATCAAGGTGGAACCGGACGTTTAGTAGGCCATATGACAGATGGCATTGGTTGTTTTAAAGCT
TTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAATTACAATAGCTGGTATTGGTGGTTTCAGGTAAAGCAGTTGCAGT
TCAAGCAGCTATGGAGGGAGTTGCGGAAATTAGATTATTTAATCGTAACAGCTCAAATTACGATAAGGTCAATTGACTTAT
CAGATAAAATTAACAAACAGTTTCAAATAAAGGTAGTCGTTGATTATCTAGAAAATAAGACAGCATTAAAGACGCCTATT
AGAAGTAGTCATTTTATATTGATGCTACTAGTTTAGGAATGAGGCCATTAGATAATTATAGTTTAAATTAACGATCCAGA
TATTTTAACACCGAATTTAGTAGTTGTGCACTT

SEQ ID NO. 1904: SAG1680 FROM THE M781 GBS TYPE III STRAIN

AAATCAGCATCCCTAGACATTATAAGCATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCT
TGTAACCAAGTCGACAACACTACTAAATTCGGTGTTAAAATTTCTGGATCGTTAATTAACTATAATTATCTAATGGCCTC
ATTCCTAAACTAGTAGCATCAATATAAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATC
AACGACTACCTTTATTTGAACTGTTTTTAAATTTTATCTGATAAGTCAATGACCTTATCGTAATTTGAGCTGTTACGAT
TAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAGTCAACTGCTTTACCTGAACCAATACCAGCTATT
GTAATTTATTTTATTTTTCAGCACTGAAACCTTGAGCTGCTAAAGCTTTAAAACAACCAATGCCATCTGTCATATGGCCTAC
TAAACGTCGGGTTCACCTTGATTACGATAGTATTTACAGCACCCTAATTTAGCTTGAGGAGATAAATCATCTAGCA
AAGGGATAACACTCTGTTTAAATGGCATTGAAACATTAACACCACGAATACTCAATGCCCTGACACCTCGAACAGCTTCT
GTTAATTTACCCTCTTCTACTTCAAATGTCAGATAGGCATAATTCATGTTTTTTCTTGAAAAGAGGTATTCCACATTAA
CGGGGATAGAGAGTGGCGTGCA

SEQ ID NO. 1905: SAG1680 FROM THE 090 GBS TYPE Ia STRAIN

GTTCGAGGTGTCAGGGCATTGGGTATTCGTGGTGTTAATGTTTCAATGCCATTTAAACAGAGTGTTATCCCTTTGCTArA
TGATTTATCTCCTCAAGCTAAATTAGTGGGTGCTGTAAATACTATCGTTAATCAAGGTGGAACCGsACGTTTAGTAGGCC
ATATGACAGATGGCATTGGTTGTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAGTTACAATAGCT
GGTATTGGTGGTTCAGGTAAAGCAGTTGCGAGTTCAAGCAGCTATGGAGGGAGTTGCGGAAATTAGATTATTTAATCGTAA
TAGCTCAAATTACGATAAGGTCAATTGACTTATCAGATAAAATTAACAAACAGTTTCAAATAAAGGTAGTCGTTGATTATC
TAGAAAATAAGACAGCATTAAAGACGCTATTAGAAGTACTGATTTTATATTGATGCTACTAGTTTAGGAATGArGCCA
TTAGATAAATTATAGTTTAAATTAACGATCCGAAAATTTAACACCCAATTTAGTAGTTGTCGACTTGGTTTACAAGCCTAA
AGAAACAGCATTGTTACGATTTGTTAGACAAAATGGAGTGAAACATGCTTATAATGGTCTAGGGATGCTGATTTATCAAG
GAGCAGA

Table 19: C mparative Sequences relating to SAG1680 (shikimate 5-dehydr genase)

SEQ ID NO. 1906: SAG1680 FROM THE A909 GBS TYPE Ia STRAIN

CCCTAGACCATTATAATCATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCA
AGTCGACAACACTACTAAATTCGGTGTTAAAATTTCTGGATCGTTAATTAACTATAATTATCTAATGGCCTCATTCTTAA
CTAGTAGCATCAATATAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTAC
CTTTATTTGAAACTGTTTTTAAATTTTATCTGATAAGTCAATGACCTTATCGTAATTTGAGCTGTTACGATTAAATAATC
TAATTTCCGCAACTCCCTCCATAGCTGCTTGAAGTCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAATTATT
TTATTTTAGCACTGAAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCTATATGGCCTACTAAACGCTCC
GGTTCACCTTGATTAACGATAGTATTTACAGCACCCTAATTTAGCTTGAGGAGATAAATCATCTAGCAAAGGGATAA
CACTCTGTTTAAATGGCATTGAAACATTAACACCACGAATACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTA
CCCTCTTCTACTTCAAATGTCAGATAGGCATAATTCATGTTTTTTCTTGAAAAGAGGTATTCCACATTAACGGGGATAG

SEQ ID NO. 1907: SAG1680 FROM THE COH1 GBS TYPE Ia STRAIN

TGCACGCCACTCTCTATCCCCGTTAATGTGGAATACCTCTTTTAAAGAAAAAACATGAATTATGCCTATCTGACATTTGA
GTAGAGAGAGGGTAAATTAACAGAAGCTGTTTCGAGGTGTGAGGGCATTGAGTATTCGTGGTGTTAATGTTTCAATGCCAT
TTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGTGCTGTAAATACT

SEQ ID NO. 1908: SAG1680 FROM THE CJB110 GBS NONTYPEABLE STRAIN

ATTCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCTAGACCATTATAAGCATGTTTCACTCCATTTT
GTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCAAGTCGACAACACTACTAAATTGGGTGTTAAAATTTCT
GGATCGTTAATTAACTATAATTATCTAATGGCCTCATTCTAAACTAGTAGCATCAATATAAAAAATGACTAGTTCTAAT
AGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTACCTTTATTTGAAACTGTTTTTAAATTTTATCTGATA
AGTCAATGACCTTATCGTAATTTGAGCTATTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAACTGCTTGAAGT
GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAAGTATTTT

SEQ ID NO. 1909: SAG1680 FROM THE CJB110 GBS NONTYPEABLE STRAIN

ACTCTCTATCCCCGTTAATGTGGAATACCTCTTTTCAAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAGAA
GAGGGTAAATTAACAGAAGCTGTTTCGAGGTGTGAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCATTTAAACA
GAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGTGCTGTAAATACTATCGTTAATCAAGGTG
GAACCGGACGTTTAGTAGGCCATATGACAGATGGCATTTGGTTGTTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAA
AATAAAATAGTTACAATAGCTGGTATTGGTG

SEQ ID NO. 1910: SAG1680 FROM THE 1169NT1 GBS TYPE V STRAIN

ATTCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCTAGACCATTATAAGCATGTTTCACTCCATTTT
GTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCAAGTCGACAACACTACTAAATTCGGTGTTAAAATTTCT
GGATCGTTAATTAACTATAATTATCTAATGGCCTCATTCTAAACTAGTAGCATCAATATAAAAAATGACTAGTTCTAAT
AGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTACCTTTATTTGAAACTGTTTTTAAATTTTATCTGATA
AGTCAATGACCTTATCGTAATTTGAGCTGTTACGAT

SEQ ID NO. 1911: SAG1680 FROM THE 1169NT1 GBS TYPE V STRAIN

ACTTCTCTATTCCCCGTTAATGTGGAATACCTCTTTTCAAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAG
AAGAGGGTAAATTAACAGAAGCTGTTTCGAGGTGTGAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCATTTAAA
CAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGTGCTGTAAATACTATCGTTAATCAAGG
TGGAACC

SEQ ID NO. 1912: SAG1680 FROM THE 18RS21 GBS TYPE II STRAIN

TCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCATCATCCCTAGACCATTATAAGCATGTTTCACTCCATTTTGT
CTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCAAGTCGACAACACTACTAAATTCGGTGTTAAAATTTCTGG
ATCGTTAATTAACTATAATTATCTAATGGCCTCATTCTAAACTAGTAGCATCAATATAAAAAATGACTAGTTCTAATAG
CGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTACCTTTATTTGAAACTGTTTTTAAATTTTATCTGATAAG
TCAATGACCTTATCGTAATTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAAC

SEQ ID NO. 1913: SAG1680 FROM THE 18RS21 GBS TYPE II STRAIN

ATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTAACAGAAGCTGTTTCGAGGTGTGAGGGCATTGGGTATTTCGTGGT
GTTAATGTTTCAATGCCATTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGTGCT
TGTAATACTATCGTTAATCAAGGTGGAACCGGACGTTTAGTAGGCCATATGACAGATGGCATTGGTTGTTTTAAAGCTT
TAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAATTACAATAGCTGGTATTGGTGGTTTCAGGTAAAGCAGTTGCAGTT
CAAGCAGCTATGGAGGGAGTTGCGG

SEQ ID NO. 1914: SAG1680 FROM THE JM9130013 GBS TYPE VIII STRAIN

CCCTAGACCATTATAAGTCATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTCTTTAGGCTTGTAACCC
AAGTCGACAACACTACTAAATTGGGTGTTAAATTTCTGGATCGTTAATTAACTATAATTATCTAATGGCCCTCATTCCCTAA
ACTAGTAGCATCAATATAAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAACGACTA
CCTTTATTTGAAACTGTTTTTAAATTTTATCTGATAAGTCAATGACCTTATCGTAATTTGAGCTATTACGATTAAATAAT
CTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAGTCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAACAT
TTTATTTTGTAGCACTGAAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCTAT

SEQ1901	-----ATCCCT
SEQ1902	-----GTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCT
SEQ1903	TGGTCTAATTGCCAATCTGCACGCCACTCTCTAT-CCCCGTTAATGTGGAATACCTCT
SEQ1904	-----AAATCAGCATCCCT
SEQ1905	-----
SEQ1906	-----CCCT
SEQ1907	-----TGCACGCCACTCTCTAT-CCCCGTTAATGTGGAATACCTCT
SEQ1908	-----ATTCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCT
SEQ1909	-----ACTCTCTAT-CCCCGTTAATGTGGAATACCTCT
SEQ1910	-----ATTCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCAGCATCCCT
SEQ1911	-----ACTTCTCTATTCCCCGTTAATGTGGAATACCTCT
SEQ1912	-----TCGTTATTAATTGAAATGCTTCTGCTCCTTGATAAATCATCATCCCT
SEQ1913	-----
SEQ1914	-----CCCT

SEQ1901	GACCATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1902	GACCATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1903	TTCAAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTA
SEQ1904	GAC-ATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1905	-----
SEQ1906	GACCATTATAAT-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1907	TT-AAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTA
SEQ1908	GACCATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1909	TTCAAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTA
SEQ1910	GACCATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1911	TTCAAGAAAAAACATGAATTATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTA
SEQ1912	GACCATTATAAG-CATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC
SEQ1913	-----ATGCCTATCTGACATTTGAAGTAGAAGAGGGTAAATTA
SEQ1914	GACCATTATAAGTCATGTTTCACTCCATTTTGTCTAACAAATCGTAACAATGCTGTTTC

SEQ1901	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1902	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1903	CAGAAGCTGTTTCGAGGTGTCAGGGCATTGAGTATTCGTGGTGTTAATGTTTCAATGCCA
SEQ1904	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1905	-----GTTTCGAGGTGTCAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCA
SEQ1906	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1907	CAGAAGCTGTTTCGAGGTGTCAGGGCATTGAGTATTCGTGGTGTTAATGTTTCAATGCCA
SEQ1908	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1909	CAGAAGCTGTTTCGAGGTGTCAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCA
SEQ1910	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1911	CAGAAGCTGTTTCGAGGTGTCAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCA
SEQ1912	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG
SEQ1913	CAGAAGCTGTTTCGAGGTGTCAGGGCATTGGGTATTTCGTGGTGTTAATGTTTCAATGCCA
SEQ1914	TTAGGCTTGTAACCAAGTC--GACAACACTAAATTCGGTGTTAAATTTCTGGATCG

Table 19: Comparative Sequences relating to SAG1680 (shikimate 5-dehydrogenase)

SEQ1901	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1902	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1903	TTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1904	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1905	TTTAAACAGAGTGTTATCCCTTTGCTARATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1906	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1907	TTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1908	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1909	TTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1910	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1911	TTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1912	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1913	TTTAAACAGAGTGTTATCCCTTTGCTAGATGATTATCTCCTCAAGCTAAATTAGTGGGT
SEQ1914	TT-AATTAACTATAATTATCT-----AATGGCCTCATTCCT-AAACTAGTAGCATCAAT
SEQ1901	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1902	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1903	CTGTAAATACTATCGTTAATCAAGGTGGAACCGGACGTTTAGTAGGCCATATGACAGAT
SEQ1904	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1905	CTGTAAATACTATCGTTAATCAAGGTGGAACCGSACGTTTAGTAGGCCATATGACAGAT
SEQ1906	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1907	CTGTAAATACT-----
SEQ1908	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1909	CTGTAAATACTATCGTTAATCAAGGTGGAACCGGACGTTTAGTAGGCCATATGACAGAT
SEQ1910	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1911	CTGTAAATACTATCGTTAATCAAGGTGGAACC-----
SEQ1912	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1913	CTGTAAATACTATCGTTAATCAAGGTGGAACCGGACGTTTAGTAGGCCATATGACAGAT
SEQ1914	TAAAAATGACTAGTTCTAATAGCGTCTTTAAATGCTGTCTTATTTTCTAGATAATCAAC
SEQ1901	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1902	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1903	GCATTGGTTGTTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAATT
SEQ1904	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1905	GCATTGGTTGTTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAGTT
SEQ1906	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1907	-----
SEQ1908	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1909	GCATTGGTTGTTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAGTT
SEQ1910	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1911	-----
SEQ1912	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1913	GCATTGGTTGTTTTAAAGCTTTAGCAGCTCAAGGTTTCAGTGCTAAAAATAAAATAATT
SEQ1914	ACTACCTTTATTTGAACTGTTTTTTAATTTTATCTGATAAGTCAATGACCTTATCGTA
SEQ1901	TTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAC
SEQ1902	TTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAC
SEQ1903	CAATAGCTGGTATTGGTGGTTCAGGTAAAGCAGTTGCAGTTCAAGCAGCTATGGAGGGA
SEQ1904	TTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAC
SEQ1905	CAATAGCTGGTATTGGTGGTTCAGGTAAAGCAGTTGCAGTTCAAGCAGCTATGGAGGGA
SEQ1906	TTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAC
SEQ1907	-----
SEQ1908	TTTGAGCTATTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAACTGCTTGAAC
SEQ1909	CAATAGCTGGTATTGGTGGT-----
SEQ1910	TTTGAGCTGTTACGAT-----
SEQ1911	-----
SEQ1912	TTTGAGCTGTTACGATTAAATAATCTAATTTCCGCAAC-----
SEQ1913	CAATAGCTGGTATTGGTGGTTCAGGTAAAGCAGTTGCAGTTCAAGCAGCTATGGAGGGA
SEQ1914	TTTGAGCTATTACGATTAAATAATCTAATTTCCGCAACTCCCTCCATAGCTGCTTGAAC

Table 19: Comparative Sequences relating to SAG1680 (shikimate 5-dehydrogenase)

SEQ1901	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAATTATTTTATTTTATAGCACT
SEQ1902	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAATTATTTTATTTTATAGCACT
SEQ1903	TTGCGGAAATTAGATTATTTAATCGTAACAGCTCAAATTACGATAAGGTCATTGACTTA
SEQ1904	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAATTATTTTATTTTATAGCACT
SEQ1905	TTGCGGAAATTAGATTATTTAATCGTAATAGCTCAAATTACGATAAGGTCATTGACTTA
SEQ1906	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAATTATTTTATTTTATAGCACT
SEQ1907	-----
SEQ1908	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAACCTATTTT-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	TTGCGG-----
SEQ1914	GCAACTGCTTTACCTGAACCACCAATACCAGCTATTGTAACCTATTTTATTTTATAGCACT
SEQ1901	AAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCATATGGCCTACTAA
SEQ1902	AAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCATATGGCCTACTAA
SEQ1903	CAGATAAAATTAAAAACAGTTTCAAATAAAGGTAGTCGTTGATTATCTAGAAAATAAG
SEQ1904	AAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCATATGGCCTACTAA
SEQ1905	CAGATAAAATTAAAAACAGTTTCAAATAAAGGTAGTCGTTGATTATCTAGAAAATAAG
SEQ1906	AAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCATATGGCCTACTAA
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	AAACCTTGAGCTGCTAAAGCTTTAAACAACCAATGCCATCTGTCAT-TABCMARAT--
SEQ1901	CGTCCGGTTCCACCTTGATTAACGATAGTATTTACAGCACCCACTAATTTAGCTTGAGG
SEQ1902	CGTCCGGTTCCACCTTGATTAACGATAGTATTTACAGCACCCACTAATTTAGCTTGAGG
SEQ1903	CAGCATTTAAAGACGCTATTAGAAGTAGTCATTTTATATTGATGCTACTAGTTTAGGA
SEQ1904	CGTCCGGTTCCACCTTGATTAACGATAGTATTTACAGCACCCACTAATTTAGCTTGAGG
SEQ1905	CAGCATTTAAAGACGCTATTAGAAGTAGTCATTTTATATTGATGCTACTAGTTTAGGA
SEQ1906	CGTCCGGTTCCACCTTGATTAACGATAGTATTTACAGCACCCACTAATTTAGCTTGAGG
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	STNCSRATNGTSASHKMATDHYDRGNAS-----
SEQ1901	GATAAATCATCTAGCAAAGGGATAAACTCTGTTTAAATGGCATTGAAACATTAACACC
SEQ1902	GATAAATCATCTAGCAAAGGGATAAACTCTGTTTAAATGGCATTGAAACATTAACACC
SEQ1903	TGAGGCCATTAGATAATTATAGTTTAATTAACGATCCAGATATTTTAAACACCGAATTTA
SEQ1904	GATAAATCATCTAGCAAAGGGATAAACTCTGTTTAAATGGCATTGAAACATTAACACC
SEQ1905	TGARGCCATTAGATAATTATAGTTTAATTAACGATCCAGAAATTTTAAACACCGAATTTA
SEQ1906	GATAAATCATCTAGCAAAGGGATAAACTCTGTTTAAATGGCATTGAAACATTAACACC
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	-----

Table 19: Comparative Sequences relating to SAG1680 (shikimate 5-dehydrogenase)

SEQ1901	CGAATACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTACCCCTCTTCTACTTC
SEQ1902	CGAATACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTACCCCTCTTCTACTTC
SEQ1903	TAGTTGTCGACTT-----
SEQ1904	CGAATACTCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTACCCCTCTTCTACTTC
SEQ1905	TAGTTGTCGACTTGGTTTACAAGCCTAAAGAAACAGCATTGTTACGATTGTTAGACAA
SEQ1906	CGAATACCCAATGCCCTGACACCTCGAACAGCTTCTGTTAATTTACCCCTCTTCTACTTC
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	-----
SEQ1901	AATGTCAGATAGGCATAATTCATGTTTTTTTCTTGAAAAGAGGTATTCCACATTAACGG
SEQ1902	AATGTCAGATAGGCATAATTCATGTTTTTTTCTTGAAAAGAGGTATTCCACATTAACGG
SEQ1903	-----
SEQ1904	AATGTCAGATAGGCATAATTCATGTTTTTTTCTTGAAAAGAGGTATTCCACATTAACGG
SEQ1905	ATGGAGTGAAACATGCTTATAATGGTCTAGGGATGCTGATTATCAAGGAGCAGA----
SEQ1906	AATGTCAGATAGGCATAATTCATGTTTTTTTCTTGAAAAGAGGTATTCCACATTAACGG
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	-----
SEQ1901	GATAGAGAGTGGCGTGCAGG--
SEQ1902	GATAGAGAGTGGCGTGCAGGA
SEQ1903	-----
SEQ1904	GATAGAGAGTGGCGTGCA---
SEQ1905	-----
SEQ1906	GATAG-----
SEQ1907	-----
SEQ1908	-----
SEQ1909	-----
SEQ1910	-----
SEQ1911	-----
SEQ1912	-----
SEQ1913	-----
SEQ1914	-----

Table 20: C mparative Sequences relating to SAG1723 (signal peptidase I)

SEQ ID NO. 2001: SAG1723 FROM THE COH1 GBS TYPE Ia STRAIN

ATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTC
ATCAAATATAAAAATGACACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGA
TAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCA
GCGAATTTACTACTGTCGTGCTAAAGGCCACTACTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGGTTCC
TTCAAAA

SEQ ID NO. 2002: SAG1680 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

TAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATA
TTGTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
AATGACACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGA
AAAATATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTTACTA
CTGTCGTGCTAAAGGCCACTATTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGTCCCTTCAAAAAATCA
ACAATTGTGGGAG

SEQ ID NO. 2003: SAG1680 FROM THE 18RS21 GBS TYPE II STRAIN

TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATATTGTA
GTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
CACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGAAAAAT
ATTGCTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACTACTGTC
GTGCTAAAGGCCACTACTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGTCCCTTCAAAAAATCAACGAT
TGTGGGAGAGGT

SEQ ID NO. 2004: SAG1680 FROM THE 2603 V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)

AAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATATT
GTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
TGACACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGAAAA
AATATTGCTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACTACT
GTCGTGCTAAAGGCCACTACTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGGT

SEQ ID NO. 2005: SAG1680 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATATTGTA
GGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGACA
CCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGAAAAATAT
TCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACTACTGTCGT
GCCTAAAGGCCACTACTATCTTGTGGTGATGACCGA

SEQ ID NO. 2006: SAG1680 FROM THE M781 GBS TYPE III STRAIN

TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATATTGTA
GTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
CACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGAAAAAT
TATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT

SEQ ID NO. 2007: SAG1680 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

TTGGTAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTC
GATATTGTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATA
TAAAAATGACACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTAC
AGGAAAAATATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTT
ACCACTGTCGTGCTAAAGGCCACTACTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGGCCCTTCAAAAA
ATCAACG

SEQ ID NO. 2008: SAG1680 FROM THE H36b GBS TYPE Ib STRAIN (REVERSE COMPLEMENT)

TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATATTGTA
GTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
CACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGAAAAAT
ATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACTACTGTC
GTGCTAAAGGCCACTACTATCTTGTGGTGATGACCGA

SEQ ID NO. 2009: SAG1680 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

TAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATA
TTGTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTCATTGGTATGCCAGGTGATGTCATCAAATATAAAAATGA
AATGACACCTTAACCTATTAACAATAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATATTTAAAAAGGATAAATTACAGGA
AAAAATATTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTTACTA
CTGTCGTGCTAAAGGCCACTATTATCTTGTGGTGATGACCGAATTGTCTCTAAAGATAGTCGTGCCGTCCGGT

Table 20: C mparative Sequences relating to SAG1723 (signal peptidase I)

SEQ ID NO. 2010: SAG1680 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)
 AAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAGTTCTCAAACAAACAAAAATCAATCGATTTCGATAT
 TGTAGTGGCTAACGAAGAAGAAGGCGGCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATAAAA
 ATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATACTAAATTATTTAAAAAGGATAAATTACAGGAA
 AAATATTTCGTATAACCCACTTTTCCAAGACCTAGCACAAAGCTCTACCGCTTTACCACTGACAGCAATGGCAGCAGCGAATTTACTAC
 TGTCGTGCCTAAAGGCCACTACTATCTTGTGGTGATGCCGAATTGTCTTAAAGATAGTCGTGCCGTCCGTCCTTCAAAAAATCAA
 CG

SEQ2001 -----
 SEQ2002 ---TAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2003 -----TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2004 -----AAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2005 -----TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2006 -----TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2007 TGGTAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2008 -----TTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2009 ---TAAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG
 SEQ2010 ----AAAGTTGACGGACACTCCATGGATCCAACCTTTAGCTGACAAGGAACAGCTAGTAG

SEQ2001 -----ATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2002 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2003 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2004 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2005 TCTCAAACAAACAAAA--TAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2006 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2007 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2008 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2009 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG
 SEQ2010 TCTCAAACAAACAAAAATCAATCGATTTCGATATTGTAGTGGCTAACGAAGAAGAAGGCG

SEQ2001 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2002 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2003 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2004 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2005 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2006 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2007 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2008 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2009 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA
 SEQ2010 GCCAAAAGAAAAAATTGTTAAACGTGTTCATTGGTATGCCAGGTGATGTCATCAAATATA

SEQ2001 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2002 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2003 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2004 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2005 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2006 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2007 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2008 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2009 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA
 SEQ2010 AAAATGACACCTTAACTATTAACAATAAAAAAACAGAAGAACCTTACCTCAAGGAATATA

SEQ2001 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2002 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2003 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2004 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2005 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2006 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2007 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2008 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2009 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA
 SEQ2010 CTAAATTATTT-AAAAAGGATAAATTACAGGAAAAATATTCGTATAACCCACTTTTCCAA

Table 20: C mparative Sequences relating to SAG1723 (signal peptidase I)

SEQ2001	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2002	GACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2003	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2004	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2005	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2006	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2007	GACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2008	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2009	GACCTAGCACAAAGCTCTACCGCTTTCACCTACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2010	GACCTAGCACAAAGCTCTACCGCTTTCACCACTGACAGCAATGGCAGCAGCGAATTTACT
SEQ2001	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2002	CTGTCGTGCCTAAAGGCCACTATTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2003	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2004	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2005	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGA-----
SEQ2006	-----
SEQ2007	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2008	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGA-----
SEQ2009	CTGTCGTGCCTAAAGGCCACTATTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2010	CTGTCGTGCCTAAAGGCCACTACTATCTTGTTGGTGATGACCGAATTGTCTCTAAAGAT
SEQ2001	GTCGTGCCGTCGGTTCCTTCAAAA-----
SEQ2002	GTCGTGCCGTCGGTCCCTTCAAAAAATCAACAATTGTGGGAG-----
SEQ2003	GTCGTGCCGTCGGTCCCTTCAAAAAATCAACGATTGTGGGAGAGGT-----
SEQ2004	GTCGTGCCGTCGGT-----
SEQ2005	-----
SEQ2006	-----
SEQ2007	GTCGTGCCGTCGGCCCCCTTCAAAAAATCAACG-----
SEQ2008	-----
SEQ2009	GTCGTGCCGTCGGT-----
SEQ2010	GTCGTGCCGTCGGTCCCTTCAAAAAATCAACGTABCMARATVSTNCSRATNGTSAGSGN
SEQ2001	----
SEQ2002	----
SEQ2003	----
SEQ2004	----
SEQ2005	----
SEQ2006	----
SEQ2007	----
SEQ2008	----
SEQ2009	----
SEQ2010	TDAS

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

SEQ ID NO. 2101: SAG0079 FROM THE 2603V/R GBS TYPE V STRAIN

AATCTTTTAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTT
GAAAAAGCGTTG

SEQ ID NO. 2102: SAG0079 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

AATCTTTTAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTT
GAAAAAGCGTTGCTAGAAGTCAAA

SEQ ID NO. 2103: SAG0079 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

TGGTAAAGGGACTCAAGCAGCTAAGATTGTTGAAGAATTTGGTGTGCGCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGC
TAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCTGATCAAGTAACAAACGGGATTGTAAA
AGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCAAGTACTATTGAACAAGCACACGCCTTAGA
TGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGATCCATCATGTCTTATAGAGCGTTTGAGTGG
TCGTATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACG
TGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTTATATTGCTCAAGGAGAACCTATTCTTGAACACTATAGTAAGCT
TGGCTTGTACAGATATTGAAGGTAATCAAGAAATAA

SEQ ID NO. 2104: SAG0079 FROM THE 18RS21 GBS TYPE II STRAIN (REVERSE COMPLEMENT)

AATCTTTTAACACCGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTT
GAAAAAGCGTTGCTAGAA

SEQ ID NO. 2105: SAG0079 FROM THE 2603V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)

AATCTTTTAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTT
GAAAAAGCGTTG

SEQ ID NO. 2106: SAG0079 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

AATCTTTTAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAATCTATTCTTGAACACTATCGAAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTT

SEQ ID NO. 2107: SAG0079 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

AATCTTTTAACACCGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTT
CCTGATGAAGTAACAAACGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGTTTTTACTTTGATGGATATCCA
CGTACTATTGAACAAGCACACGCCTTAGATGCTACGCTTGAAGAACTAGGACTACGCTTAGATGGTGTATTAAATATTAAGTGGAT
CCATCATGTCTTATAGAGCGTTTGAGTGGTCTGATTATCAATCGTAAAGTGGTGAAGTCTTCCACAAAGTGTTCACCCACCAGTA
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTGCGTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATAG

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

SEQ ID NO. 2108: SAG0079 FROM THE COH1 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATTGTTGAAGAATTTGGTGTGCTCACATCTCAA
CAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCCAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTC
CTGATGAAGTAACAAACGGGATTGTAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAGGTTTTTACTTGTATGGATATCCAC
GTACTATTGAGCAAGCACACGCCCTTAGATGCTACGCTTGAAGAAGTACGCTTAGATGGTGTATTAAATATTAAAGTGGATC
CAACATGCCTTATAGAGCGTTTGAAGTGGCCGTATTATCAATCGTAAAGTGGTGAAGCTTCCACAAAGTGTTCACCCACCAGTAG
ATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAG
AACCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTTG
AAAAAGCGTTGCTAG

SEQ ID NO. 2109: SAG0079 FROM THE H36b GBS TRYP ID STRAIN (REVERSE COMPLEMENT)

CAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTC
CTGATGAAGTAACAAACGGGATTGTAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAGGTTTTTACTTGTATGGATATCCAC
GTACTATTGAACAAGCACACGCCCTTAGATGCTACGCTTGAAGAAGTACGCTTAGATGGTGTATTAAATATTAAAGTGGATC
CATCATGCTTATAGAGCGTTTGAAGTGGTCTATTATCAATCGTAAAGTGGTGAAGCTTCCACAAAGTGTTCACCCACCAGTAG
ATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAG
AATCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTTG
AAAAAGCGTTGCT

SEQ ID NO. 2110: SAG0079 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)

AATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTC
CCTGATGAAGTAACAAACGGGATTGTAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAGGTTTTTACTTGTATGGATATCCAC
CGTACTATTGAACAAGCACACGCCCTTAGATGCTACGCTTGAAGAAGTACGCTTAGATGGTGTATTAAATATTAAAGTGGATC
CCATCATGCTTATAGAGCGTTTGAAGTGGTCTATTATCAATCGTAAAGTGGTGAAGCTTCCACAAAGTGTTCACCCACCAGTAG
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGA
GAACCTATTCTTGAACACTATAAAAGCTTGGTCTTGTACAGATATTGAAGGTAATCA

SEQ ID NO. 2111: SAG0079 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

CTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATTGTTGAAGAATTTGGTGTGCTCACATCTCAACA
GGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCCAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCT
GATGAAGTAACAAACGGGATTGTAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAGGTTTTTACTTGTATGGATATCCACGT
ACTATTGAGCAAGCACACGCCCTTAGATGCTACGCTTGAAGAAGTACGCTTAGATGGTGTATTAAATATTAAAGTGGATCCA
ACATGCCTTATAGAGCGTTTGAAGTGGTCTATTATCAATCGTAAAGTGGTGAAGCTTCCACAAAGTGTTCACCCACCAGTAGAT
TATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAA
CCTATTCTTGAACACTATCGTAAGCTTGGTCTTGTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTTGCAGATGTTGAA
AAAGCGTTGCTAGAAGTCAAA

SEQ ID NO. 2112: SAG0079 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

AATCTTTTAAATTACGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATTGTTGAAGAATTTGGTGTGCTCACATCTCA
ACAGGGGATATGTTCCGCGCCGCAATGGCTAATCAAACCCAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTC
CCTGATGAAGTAACAAACGGGATTGTAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAGGTTTTTACTTGTATGGATATCCAC
CGTACTATTGAGCAAGCACACGCCCTTAGATGCTACGCTTGAAGAAGTACGCTTAGATGGTGTATTAAATATTAAAGTGGATC
CCAACATGCCTTATAGAGCGTTTGAAGTGGCCGTATTATCAATCGTAAAGTGGTGAAGCTTCCACAAAGTGTTCACCCACCAGTAG
GATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCTGAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAA

SEQ2101	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2102	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2103	-----TGGTAAAGGGACTCAAGCAGCTAAGATTGTT
SEQ2104	ATCTTTTAAACACGGGTTTCGCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2105	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2106	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2107	ATCTTTTAAACACGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2108	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2109	-----
SEQ2110	ATCTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATCGTT
SEQ2111	--CTTTTAAATTATGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATGTT
SEQ2112	ATCTTTTAAATTACGGGTTTGCCTGGTGTGCTGGTAAAGGTAAGCTCAAGCAGCTAAGATGTT

SEQ2101	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2102	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2103	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2104	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2105	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2106	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2107	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2108	AAGAATTTGGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

SEQ2109	-----CAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2110	AAGAATTTGGTGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2111	AAGAATTTGGTGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2112	AAGAATTTGGTGTGTGCTCACATCTCAACAGGGGATATGTTCCGCGCCGCAATGGCTAAT
SEQ2101	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2102	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2103	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2104	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2105	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2106	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2107	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2108	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2109	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2110	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2111	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2112	CAAACCGAAATGGGACGTTTAGCTAAAAGTTATATTGATAAAGGTGAATTGGTTCCTGAT
SEQ2101	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2102	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2103	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2104	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2105	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2106	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2107	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2108	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2109	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2110	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2111	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2112	AAGTAACAAACCGGGATTGTAAAAGAGCGCTTAGCTGAGGATGATATCGCAGAAAAAGGT
SEQ2101	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2102	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2103	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2104	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2105	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2106	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2107	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2108	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2109	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2110	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2111	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2112	TTTTTACTTGATGGATATCCACGTAATTTGAACAAGCACACGCCTTAGATGCTACGCTT
SEQ2101	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2102	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2103	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2104	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2105	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2106	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2107	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2108	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2109	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2110	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2111	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2112	GAAGAACTAGGACTACGCTTAGATGGTGTATTATTAATATTAAAGTGGATCCATCATGTCTT
SEQ2101	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2102	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2103	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2104	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2105	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2106	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2107	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2108	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2109	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2110	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2111	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG
SEQ2112	ATAGAGCGTTTGAGTGGTTCGTATTATCAATCGTAAAACCTGGTGAAACTTTCCACAAAGTG

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

SEQ2101	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2102	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2103	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2104	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2105	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2106	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2107	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2108	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2109	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2110	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2111	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2112	TTCAACCCACCAGTAGATTATAAAGAAGAAGATTACTATCAACGTGAAGATGATAAGCCT
SEQ2101	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2102	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2103	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2104	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2105	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2106	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2107	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2108	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2109	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2110	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2111	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAAGGAGAACCTATTCTTGAACAC
SEQ2112	GAAACTGTCAAACGTCGCTTGGACGTTAATATTGCTCAATABC MARATVSTNCSR---AT
SEQ2101	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2102	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2103	ATAGTAAGCTTGGCCTTGTTACAGATATTGAAGGTAATCAAGAAATAA-----
SEQ2104	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2105	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2106	ATCGAAAGCTTGGTCTTGTTACAGATATTGAAGGTAA-----
SEQ2107	ATAG-----
SEQ2108	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2109	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2110	ATAAAAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCA-----
SEQ2111	ATCGTAAGCTTGGTCTTGTTACAGATATTGAAGGTAATCAAGAAATAACAGAAGTTTTT
SEQ2112	GTSAGADNYATKNAS-----
SEQ2101	CAGATGTTGAAAAAGCGTTG-----
SEQ2102	CAGATGTTGAAAAAGCGTTGCTAGAACTCAA
SEQ2103	-----
SEQ2104	CAGATGTTGAAAAAGCGTTGCTAGAA-----
SEQ2105	CAGATGTTGAAAAAGCGTTG-----
SEQ2106	-----
SEQ2107	-----
SEQ2108	CAGATGTTGAAAAAGCGTTGCTAG-----
SEQ2109	CAGATGTTGAAAAAGCGTTGCT-----
SEQ2110	-----
SEQ2111	CAGATGTTGAAAAAGCGTTGCTAGAACTCAA
SEQ2112	-----

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

>SEQ ID NO 2150:090 frame: 1

NLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEH
YRKLGLVTDIEGNQEITEVFADVEKALLELK

>SEQ ID NO 2151:114_1169NT frame: 2

GKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPDQVTNGIVKER
LAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC LIERLSGRIIN
RKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVHIAQGEPILEHYSKLG LVTDI
EGNQEI

>SEQ ID NO 2152: 114_18RS21 frame: 1

NLLTTGSPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEH
YRKLGLVTDIEGNQEITEVFADVEKALLE

>SEQ ID NO 2153: 114_2603 frame: 1

NLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEH
YRKLGLVTDIEGNQEITEVFADVEKAL

>SEQ ID NO 2154: 114_A909 frame: 1

NLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGESILEH
YRKLGLVTDIEG

>SEQ ID NO 2155:114_A909 frame: 1

NLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGESILEH
YRKLGLVTDIEG

>SEQ ID NO 2156: 114_CJB110 frame: 1

NLLTTGLLGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEH
Y

>SEQ ID NO 2157: 114_COH1 frame: 3

LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTQMGRLAKSYIDKGELVPDE
VTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPTCLI
ERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEHY
RKLGLVTDIEGNQEITEVFADVEKALL

>SEQ ID NO 2158: 114_H36B frame: 3

GDMFRAAMANQTEMGR LAKSYIDKGELVPDEVTNGIVKERLAEDDIAEKGFLLDGYPR TI
EQAHALDATLEELGLRLDGVINIKVDPSC LIERLSGRIINRKTGETFHKVFNPPVDYKEE
DYYQREDDKPETVKRRLDVNIAQGESILEHYRKLGLVTDIEGNQEITEVFADVEKAL

>SEQ ID NO 2159: 114_JM9130013 frame: 1

NLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGR LAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPSC L
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEH
YKKLGLVTDIEGN

>SEQ ID NO 2160:114_M732 frame: 1

LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTQMGRLAKSYIDKGELVPDE
VTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPTCLI
ERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPILEHY
RKLGLVTDIEGNQEITEVFADVEKALLELK

>SEQ ID NO 2161: 114_M781 frame: 1

NLLITGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTQMGRLAKSYIDKGELVPD
EVTNGIVKERLAEDDIAEKGFLLDGYPR TIEQAHALDATLEELGLRLDGVINIKVDPTCL
IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIAQ

Table 21: Comparative Sequences relating to SAG0079 (adenylate kinase)

SEQ2150	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2151	-----GKGTTAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2152	LLTTGSPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2153	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2154	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2155	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2156	LLTTGLLGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2157	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2158	-----GDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2159	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2160	LLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2161	LLITGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRLAKSYIDKGELVPD
SEQ2150	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2151	QVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2152	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2153	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2154	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2155	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2156	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2157	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPTCL
SEQ2158	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2159	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPSCL
SEQ2160	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPTCL
SEQ2161	EVTNGIVKERLAEDDIAEKGFLLDGYPRTEQAHALDATLEELGLRLDGVINIKVDPTCL
SEQ2150	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2151	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVHIAQGEPILEH
SEQ2152	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2153	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2154	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGESILEH
SEQ2155	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGESILEH
SEQ2156	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2157	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2158	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGESILEH
SEQ2159	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2160	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQAQGEPILEH
SEQ2161	IERLSGRIINRKTGETFHKVFNPPVDYKEEDYYQREDDKPETVKRRLDVNIQA-----
SEQ2150	RKLGLVTDIEGNQEITEVFADVEKALLELK
SEQ2151	SKLGLVTDIEGNQEI-----
SEQ2152	RKLGLVTDIEGNQEITEVFADVEKALLE--
SEQ2153	RKLGLVTDIEGNQEITEVFADVEKAL----
SEQ2154	RKLGLVTDIEG-----
SEQ2155	RKLGLVTDIEG-----
SEQ2156	-----
SEQ2157	RKLGLVTDIEGNQEITEVFADVEKALL---
SEQ2158	RKLGLVTDIEGNQEITEVFADVEKAL----
SEQ2159	KKLGLVTDIEGN-----
SEQ2160	RKLGLVTDIEGNQEITEVFADVEKALLELK
SEQ2161	-----

**Table 22: Comparative Sequences relating t SAG0093
(D-alanyl-D-alanine carboxypeptidase family protein)**

SEQ ID NO. 2201: SAG0093 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

AAGCCTAACAGTCAACAATCATCATCTCAAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATAAGAAATTA
CAATTACCAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTT
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGAA
CATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCTAAT
TTGACGAGGGGACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2202: SAG0093 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

AAGCCTAACAGTCAACAATCATCACCTCAAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATAAGAAATTA
CGATTACCAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTG
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGAA
CATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCTAAT
TTGACGAGGGGACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAAACATCGTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2203: SAG0093 FROM THE 18RS21 GBS TYPE II STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATAAGAAATTA
CAATTACCAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTT
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGAA
CATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCTAAT
TTGACGAGGGGACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2204: SAG0093 FROM THE 2603V/R GBS TYPE V STRAIN

ACAGTCAACAATCATCATCTCAAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATAAGAAATTACAATTAC
CAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTTCTGTTG
AAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGAACATTTAA
TTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCTAATTTGACGA
GGGGACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGATGAGTA
CTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACCGTTTC
CGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATATATGGCCA
AACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAAACCCAGCTTTCTTGACAA

SEQ ID NO. 2205: SAG0093 FROM THE A909 GBS TYPE Ia STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAAGTTGAGGAATGAGGATATAAAAAAGACATCCTCTCAAAAAAGAAATAAGAAATTA
CGATTACCAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTG
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGAA
CATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAAATGACTAGTAACCTAAT
TTGACGAAGGAACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2206: SAG0093 FROM THE CJB110 GBS NONTYPEABLE STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATAAGAAATTT
ACAATTACCAGCTGTATCATCAAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGT
TCCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTAGAGGCTGCTAGAGCAATTGATTCACGAGA
ACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAAATGACTAGTAACCTAA
TTTGACGAGGGGACAAGCAGAAAAGTTGGTAAAACTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGA
TATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTT
ACGGTTTCCGGATGGTAAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATA
TATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

**Table 22: Comparative Sequences relating to SAG0093
(D-alanyl-D-alanine carboxypeptidase family protein)**

SEQ ID NO. 2207: SAG0093 FROM THE COH1 GBS TYPE III STRAIN

CCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACATCCTCTCAAAAAAGAAATTAAGAAATTACG
ATTACCAGCTGTATCATCAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCC
TGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTTAGAGGCTGCTAGAGCAATTGATTACAGAGAACA
TTTAATTTTCGGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCCTAATTT
GACGAGGGGACAAGCAGAAAAGTTGTAATAAATTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGATAT
GAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAGATAGCTCCACAATATGGTTTTGTCTTACG
GTTTCCGGATGGTAAACAGCAGAAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATATAT
GGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAAACCCAGCTTTCTGTACAA

SEQ ID NO. 2208: SAG0093 FROM THE H36b GBS TYPE Ib STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACATCCTCTCAAAAAAGAAATTAAGAAATTA
CGATTACCAGCTGTATCATCAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTG
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTTAGAGGCTGCTAGAGCAATTGATTACAGAGAA
CATTTAATTTTCGGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAWGAATGACTAGTAACCCTAAT
TTGACGAAGGAACAAGCAGAAAAGTTGTAATAAATTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2209: SAG0093 FROM THE JM9130013 GBS TYPE VIII STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATATCCTCTCAAAAAAGAAATTAAGAAATTA
CAATTACCAGCTGTATCATCAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTT
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTTAGAGGCTGCTAGAGCAATTGATTACAGAGAA
CATTTAATTTTCGGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCCTAAT
TTGACGAGGGGACAAGCAGAAAAGTTGTAATAAATTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGCCAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ ID NO. 2210: SAG0093 FROM THE M732 GBS TYPE III STRAIN

AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACATCCTCTCAAAAAAGAAATTAAGAAATTAC
GATTACCAGCTGTATCATCAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTG
CTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTTAGAGGCTGCTAGAGCAATTGATTACAGAGAA
ATTTAATTTTCGGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCCTAAT
TGACGAGGGGACAAGCAGAAAAGTTGTAATAAATTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGATA
TGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAGATAGCTCCACAATATGGTTTTGTCTTAC
GGTTTCCGGATGGTAAACAGCAGAAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATATA
TGGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAAACCCAGCTTTCTT

SEQ ID NO. 2211: SAG0093 FROM THE M781 GBS TYPE III STRAIN

AAGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACATCCTCTCAAAAAAGAAATTAAGAAATTA
CGATTACCAGCTGTATCATCAAAGATTGGAACCTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTG
CCTGTTGAAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTTAGAGGCTGCTAGAGCAATTGATTACAGAGAA
CATTTAATTTTCGGGTTATCGTAGTGTTCCTATCAGGAGAAGTTGTTCAATTCCTATGTTACTCAAGAGATGACTAGTAACCCTAAT
TTGACGAGGGGACAAGCAGAAAAGTTGTAATAAATTTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGACTGGATTAGCGATGGAT
ATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAGTCAGTCAGTTGAAAAGATAGCTCCACAATATGGTTTTGTCTTA
CGGTTTCCGGATGGTAAACAGCAGAAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGTCTGCAAAATAT
ATGGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGGAGAATAACCAA

SEQ2201	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2202	AGCCTAACAGTCAACAATCATCACCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2203	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2204	-----ACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2205	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACA
SEQ2206	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2207	--CCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACA
SEQ2208	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACA
SEQ2209	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGATA
SEQ2210	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACA
SEQ2211	AGCCTAACAGTCAACAATCATCATCTCAAAGTTGAGGAATGAGGATATAAAAAAGACA

**Tabl 22: C mparative Sequences relating to SAG0093
(D-alanyl-D-alanine carboxypeptidase family protein)**

SEQ2201	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACAATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2202	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2203	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACAATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2204	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACAATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2205	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2206	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2207	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2208	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2209	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACAATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2210	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2211	TCCTCTCAAAAAAGAAAT-AAGAAATT-ACGATTACCAGCTGTATCATCAAAAGATTGGA
SEQ2201	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTTCCTG
SEQ2202	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2203	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTTCCTG
SEQ2204	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTTCCTG
SEQ2205	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2206	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2207	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2208	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2209	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTTCCTG
SEQ2210	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2211	ACTTGATTTTGGTCAATCGTGACCATAAACATGAAGAATTAAGTCCAGATGTGGTGCCTG
SEQ2201	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2202	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2203	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2204	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2205	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2206	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2207	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2208	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2209	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2210	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2211	TTGAAATATTTATTTGGATAAACGTATTACGAAGCAAGCTACTCAGTTTTATAGAGGCTG
SEQ2201	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2202	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2203	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2204	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2205	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2206	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2207	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2208	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2209	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2210	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2211	CTAGAGCAATTGATTACAGGAGAACATTTAATTTCCGGTTATCGTAGTGTTCCTATCAGG
SEQ2201	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2202	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2203	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2204	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2205	AGAAGTTGTTCAATTCTTATGTTACTCAAGAAATGACTAGTAACCCCTAATTTGACGAAGG
SEQ2206	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2207	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2208	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAAGG
SEQ2209	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2210	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG
SEQ2211	AGAAGTTGTTCAATTCTTATGTTACTCAAGAGATGACTAGTAACCCCTAATTTGACGAGGG

**Table 22: Comparative Sequences relating to SAG0093
(D-alanyl-D-alanine carboxypeptidase family protein)**

SEQ2201	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2202	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2203	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2204	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2205	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2206	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2207	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2208	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2209	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2210	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2211	ACAAGCAGAAAAGTTGGTAAAAAATTACTCTCAGCCTGCAGGTGCTAGTGAACACCAGA
SEQ2201	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2202	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2203	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2204	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2205	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2206	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2207	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2208	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2209	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2210	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2211	CTGGATTAGCGATGGATATGAGTACTGTAGATTCTTTGAATGAGAGCGATCCTAGAGTAG
SEQ2201	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2202	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2203	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2204	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2205	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2206	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2207	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2208	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2209	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2210	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2211	TCAGTCAGTTGAAAAAGATAGCTCCACAATATGGTTTTGTCTTACGGTTTCCGGATGGTA
SEQ2201	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2202	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2203	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2204	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2205	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2206	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2207	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2208	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2209	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2210	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2211	AAACAGCAGAAACAGGGGTAGGTTATGAAGATTGGCATTACCGCTATGTTGGGGTAGAGT
SEQ2201	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2202	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2203	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2204	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2205	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2206	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2207	CTGCAAAATATATGGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2208	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2209	CTGCAAAATATATGGCCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2210	CTGCAAAATATATGGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG
SEQ2211	CTGCAAAATATATGGTCAAACATCATTTAACATTAGAAGAATACATAACTTTATTAAAGG

**Table 22: Comparative Sequences relating to SAG0093
(D-alanyl-D-alanine carboxypeptidase family protein)**

SEQ2201	AGAATAACCAA-----
SEQ2202	AGAATAACCAA-----
SEQ2203	AGAATAACCAA-----
SEQ2204	AGAATAACCAAAACCCAGCTTTCTTGTACAA-----
SEQ2205	AGAATAACCAA-----
SEQ2206	AGAATAACCAA-----
SEQ2207	AGAATAACCAAAACCCAGCTTTCTTGTACAA-----
SEQ2208	AGAATAACCAA-----
SEQ2209	AGAATAACCAA-----
SEQ2210	AGAATAACCAAAACCCAGCTTTCTT-----
SEQ2211	AGAATAACCAATABC MARATVSTNCSRATNGTSAGDAANYDAANNCARBX YTDASAMYRT

>SEQ ID NO 2250: 18_090 frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLQLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ

>SEQ ID NO 2251: 18_1169NT frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLRLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAEHRITLEEYITLLKENNQ

>SEQ ID NO 2252: 18_18RS21 frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLQLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ

>SEQ ID NO 2253: 18_2603 frame: 3

SQSSSQKL RNEDIKKTSSQKR NKKLQLPAVSSK DWNLILVNRD HKHEELSPDVVPVENT
YLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRGQAE
KLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGKTAE
TGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQNP AELY

>SEQ ID NO 2254: 18_A909 frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLRLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTKE
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ

>SEQ ID NO 2255: 18_CJB110 frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKFTITSCI KRLELDFGQS

>SEQ ID NO 2256: 18_COH1 frame: 1

PNSQQSSSQKL RNEDIKKTSSQKR N

>SEQ ID NO 2257: 18_H36B frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLRLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTXEMTSNPNLTKE
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ

>SEQ ID NO 2258: 18_JM9130013 frame: 1

KPNSQQSSSQKL RNEDIKKTSSQKR NKKLQLPAVSSK DWNLILVNRD HKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
QAEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGK
TAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ

>SEQ ID NO 2259: 18_M732 frame: 3

PNSQQSSSQKL RNEDIKKTSSQKR NKKLRLPAVSSK DWNLILVNRD HKHEELSPDVVPVE
NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRGQ
AEKLVKTY SQPAGASEHQTGLAMD MSTVDSL NESDPRVVSQ LKKIAPQYGFVLRFPDGKT
AETGVGYEDWHYRYVGVESAKYMKHHLTLEEYITLLKENNQNP AF

**Tabl 22: Comparative Sequences relating to SAG0095
(D-alanyl-D-alanine carboxypeptidase family protein)**

>SEQ ID NO 2260: 18_M781 frame: 1

KPNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
ENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
QAEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
TAETGVGYEDWHYRYVGVESAKYMKHHLTLEEYITLLKENNQ

SEQ2250 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2251 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2252 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2253 --SQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2254 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2255 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2256 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2257 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2258 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2259 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV
SEQ2260 PNSQQSSSQKLARNEDIKKTSSQKRNNKLRLEAVSSKDWNLILVNRDHKHEELSPDVVPV

SEQ2250 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2251 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2252 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2253 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2254 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2255 -----
SEQ2256 -----
SEQ2257 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTXEMTSNPNLTRG
SEQ2258 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2259 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG
SEQ2260 NIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEMTSNPNLTRG

SEQ2250 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2251 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2252 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2253 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2254 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2255 -----
SEQ2256 -----
SEQ2257 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2258 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2259 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK
SEQ2260 AEKLVKTYSQPAGASEHOTGLAMDMSVDSLNESDPRVVSQKKIAPQYGFVLRFPDGGK

SEQ2250 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ-----
SEQ2251 AETGVGYEDWHYRYVGVESAKYMAEHRLTLEEYITLLKENNQ-----
SEQ2252 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ-----
SEQ2253 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQNONPAFLY-----
SEQ2254 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ-----
SEQ2255 -----
SEQ2256 -----
SEQ2257 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ-----
SEQ2258 AETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ-----
SEQ2259 AETGVGYEDWHYRYVGVESAKYMKHHLTLEEYITLLKENNQNONPAF-----
SEQ2260 AETGVGYEDWHYRYVGVESAKYMKHHLTLEEYITLLKENNQTABLECMPARATIVESE

SEQ2250 -----
SEQ2251 -----
SEQ2252 -----
SEQ2253 -----
SEQ2254 -----
SEQ2255 -----
SEQ2256 -----
SEQ2257 -----
SEQ2258 -----
SEQ2259 -----
SEQ2260 ENCESRELATINGTSAGDALANYLDALANINECARBYPEPTIDASEFAMILYPRTEIN

**Table 23: Comparative Sequences relating to SAG0163
(competence protein CglA)**

SEQ ID NO. 2301: SAG0163 FROM THE 090 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
 GGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTCCTCCAAAGGTGATTGTTATGAAGTCTATATGCGTATTGATGATGAAAGGCG
 GTTTATTGATGTTTTTGAAGTTAATAGGATGGCTAGTCTTATTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAG
 ACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTGG
 TCAAGAATCTTTAGTTATTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTTGGTTTGATAATATAAGCAAATGAAGGAAGT
 ACTGGGTACAAGAGGGCTATATCTTTTTTCCGGCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATT
 TAAAAATAAGCAAATTATCAGGATTTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAACCTCAATTGAATGAGGATATTGG
 AATGACTTATGATGCTTTAATCAAACGTCTTTACGGCATCGTCCAGATATTTTAAATATCGGAGAGATTAGAGATCAAGCGACGGC
 CCGTGTCTGTTATTCGTGCAAGTTTAAACGGGAGTGATGGTTTTTCTACTATTTCATGCTAAAAGTATTTCCGGAGTCTATGATAGGCT
 TATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAAT
 TGACTTTGAGACAGGTAACCTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAAGGACATATCAG
 TAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGTCCAACCTTTT

SEQ ID NO. 2302: SAG0163 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)
 GGTGATTGTTATGAAACCTCTACTATTGCGTATTTGATGATGAAAGGCGGTTTATTGATGTTTTTGAAGTTAATAGGATGGCTAGTC
 TTATTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAG
 AGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTTCAGGTC
 ATCAGGACTTAAATATTTGGTTTGATAATATAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCGGCCCTG
 TGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCAGGATTTGAAGATCCGGTAG
 AATCAAGAATGACAAGATGTTACAACCTCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
 ATCGTCCAGATATTTAATTATCGGAGAGATTAGAGATCAAGCGACGGCTCGTGTCTGTTATTCGTGCAAGTTTAAACGGGAGTGATGG
 TTTTTTCTACTATTTCATGCTAAAAGTATTTCCCGGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATA
 GTCTAAAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTTAAAAAACACTCATCAG
 ACAAGTGAATAGACAAGTGGATATCTTGGCTGAAGAAGGATATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAG
 AAACAACGGAAAGTAGTCCAACCTTTT

SEQ ID NO. 2303: SAG0163 FROM THE 18RS21 GBS TYPE II STRAIN (REVERSE COMPLEMENT)
 GTTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTCCCAAGGTGATTGTTAT
 GAAGTCTATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTTGAAGTTAATAGGATGGCTAGTCTTATTAGTCACTTTAAA
 TTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGGAAGACTGGTTTCA
 TACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATAT
 TGGTTTGATAATATAAAGCAAATGAAGGAAGTACTGGGTATAAGAGGGCTATATCTTTTTTCCGGCCCTGTGGGGAGTGGTAAACAA
 ACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCAGGATTTGAAGATCCGGTAGAAATCAAGAATGACAAG
 ATGTTACAACCTCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGCATCGTCCAGATATTTTA
 ATTATCGGAGAGATTAGAGATCAAGCGACGGCCCGTGTCTGTTATTCGTGCAAGTTTAAACGGGAGTGATGGTTTTTCTACTATTATC
 GCTAAAAGTATTTCCCGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCA
 TATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAA
 GTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGT
 CCAACCTTTT

SEQ ID NO. 2304: SAG0163 FROM THE 2603 V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)
 GATATTTATATCATTCCCAAGGTGATTGTTATGAAGTCTATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTTGAAGTTT
 AATAGGATGGCTAGTCTTATTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGT
 GACTATGAAGTGTGAGAGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGT
 ATTTTGTATTTCAGGTCATCAGGACTTAAATATTTGGTTTGATAATATAAGCAAATGAAGGAAGTACTGGGTATAAGAGGGCTATAT
 CTTTTTCCGGCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCAGC
 ATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAACCTCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATC
 AAACGTCTTTACGGCATCGTCCAGATATTTAATTATCGGAGAGATTAGAGATCAAGCGACGGCCCGTGTCTGTTATTCGTGCAAGT
 TTAACGGGAGTGATGGTTTTTCTACTATTTCATGCTAAAAGTATTTCCCGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTAT
 CAAGAGTTAGAAAATAGTCTAAAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTT
 AAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTGGCA
 AAAATTATCCCTCAAGAAACAACGGAAAGTAGTCCAACCTTTT

**Table 23: Comparative Sequences relating to SAG0163
(competence protein CglA)**

SEQ ID NO. 2305: SAG0163 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

GTTCATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTCCCAAAGGTGATTGTTAT
GAACTCTATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTGAGTTTAAATAGGATGGCTAGTCTTATTAGTCACTTTAAA
TTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGAAGACTGGTTTCA
TTACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTCAGGTCATCAGGACTTAAATAT
TGGTTTGATAATATAAAGCAAATGAAGGAAGTACTGGGTATAAGAGGGCTATATCTTTTTCCGGCCCTGTGGGGAGTGGTAAAAACA
ACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAG
ATGTTACAACCTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAAGTGTCTTTACGGCATCGTCCAGATATTTTA
ATTATCGGAGAGATTAGAGATCAAGCGACGGCCCGTGTCTTATTGCTGCAAGTTTAAACGGGAGTGATGGTTTTTCTACTATTTCAT
GCTAAAAGTATTTCCCGAGTCTATGATAGGCTTATAGAAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCA
TATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTTAAAAAACACTCATCAGACAAGTGAATAGACAA
GTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGT
CCAACTTTT

SEQ ID NO. 2306: SAG0163 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

GTTCATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTCCCAAAGGTGATTGTTAT
GAACTCTATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTGAGTTTAAATAGGATGGCTAGTCTTATTAGTCACTTTAAA
TTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGAAGACTGGTTTCA
TTACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTCAGGTCATCAGGACTTAAATAT
TGGTTTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAGAGGGCTATATCTTTTTCCGGCCCTGTGGGGAGTGGTAAAAACA
ACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAG
ATGTTACAACCTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAAGTGTCTTTACGGCATCGTCCAGATATTTTA
ATTATCGGAGAGATTAGAGATCAAGCGACGGCCCGTGTCTTATTGCTGCAAGTTTAAACGGGAGTGATGGTTTTTCTACTATTTCAT
GCTAAAAGTATTTCCCGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCA
TATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAACTTTAAAAAACACTCATCAGACAAGTGAATAGACAA
GTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGT
CCAACTTTT

SEQ ID NO. 2307: SAG0163 FROM THE COH1 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

AGGTGATTGTTATGAAATTTCTATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTGAGTTTAAATAGGATGGCTAGTCTTA
TTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGG
GAAGACTGGTTTCAATTACGACTATCAAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTACTTTGTATTTCAGGTCATC
AGGACTTAAATATTTGGTTTGATAATATAAAGTAAATGAAGGAAGTACTGTGTGCAAGAGGGCTATATCTTTTTCCGGCCCTGTGG
GGAGTGGTAAAAACAACCTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAA
TCAAGAATGACAAGATGTTACAACCTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAAGTGTCTTTACGGCATC
GTCCAGATATTTTAAATTATCGGAGAGATTAGAGATCAAGCGACGGCCCGTGTCTTATTGCTGCAAGTTTAAACGGGAGTAATGGTTT
TTTCTACTATTTCATGCTAAAAGTATTTCCCGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTC
TAAAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAAGTAACTTTAAAAAACACTCATCAGACA
AGTGAATAGACAAGTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAA
CAACGGAAAGTAGTCCAACTTTT

SEQ ID NO. 2308: SAG0163 FROM THE H36b GBS TYPE Ib STRAIN (REVERSE COMPLEMENT)

TCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTCCCAAAGGTGATTGTTATGAAGTCT
TATATGCGTATTGATGATGAAAGGCGGTTTATTGATGTTTTGAGTTTAAATAGGATGGCTAGTCTTATTAGTCACTTTAAATTTGTG
GCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGAAGACTGGTTTCAATTACGA
CTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTCAGGTCATCAGGACTTAAATATTTGGTTT
GATAATATAAAGCAAATGAAGGAAGTACTGGGTATAAGAGGGCTATATCTTTTTCCGGCCCTGTGGGGAGTGGTAAAAACAACCTCTC
ATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTA
CAACTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAAGTGTCTTTACGGCATCGTCCAGATATTTTAAATATC
GGAGAGAAATAGAGATCAAGCGACGGCCCGTGTCTTATTGCTGCAAGTTTAAACGGGAGTGATGTTTTTCTACTATTTCATGCTAA
AAGTATTTCCCGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCATATCA
ACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTTAAAAAACACTCATCAGACAAGTGAATAGACAAGTGA
TATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGTCCAAC
TTTT

**Table 23: Comparative Sequences relating to SAG0
(competence protein Cg1A)**

SEQ ID NO. 2309: SAG0163 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)
 GTTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATTTATATCATTTCCCAAAGGTGATTGTTAT
 GAACCTCTATATGCGTATTGATGATGAAAGCGGGTTTATTGATGTTTTTGAGTTTAAATAGGATGGCTAGTCTTATTAGTCACTTTAAA
 TTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGGAAGACTGGTTTCA
 TTACGACTATCGAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTATTTTGTATTTCAGGTCTATCAGGACTTAAAATAT
 TGGTTTGATAATATAAAGCAAATGAAGGAAGTACTGGGTATAAGAGGGCTATATCTTTTCCGGCCCTGTGGGGAGTGGTAAAACA
 ACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAG
 ATGTTACAACCTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTGTCTTTACGGCATCGTCCAGATATTTTA
 ATTATCGGAGAGATTAGAGATCAAGCGACGGCCCCGTGCTGTTATTCGTGCAAGTTTAAACGGGAGTGATGGTTTTTCTACTATTTCAT
 GCTAAAAGTATTTCCCGGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCA
 TATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGTAATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAA
 GTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGT
 CCAACTTTT

SEQ ID NO. 2310: SAG0163 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
 TGACTTGTTATGAAACTCTATATGCGTATTGATGATGAAAAGCGGTTTATTGATGTTTTTGAGTTTAAATAGGATGGCTAGTCTTA
 TTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAGACGAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGG
 GAAGACTGGTTTCATTACGACTATCAAGTGTGGGAGATTATCGTGGTCAAGAATCTTTAGTTATTCGTACTTTGTATTTCAGGTCTATC
 AGGACTTAAAATATTTGGTTTGATAATATAAAGTAAATGAAGGAAGTACTGTGTGCAAGAGGGCTATATCTTTTTCCGGCCCTGTGG
 GGAGTGGTAAACAACCTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAA
 TCAGAATGACAAGATGTTACAACCTCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTGTCTTTACGGCATC
 GTCCAGATATTTTAATTATCGGAGAGATTAGAGATCAAGCGACGGCCCCGTGCTGTTATTCGTGCAAGTTTAAACGGGAGTAATGGTTT
 TTTCTACTATTTCATGCTAAAAGTATTTCCCGGAGTCTATGATAGGCTTATAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTC
 TAAAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAAGTAACTTTAAAAAACACTCATCAGACA
 AGTGAATAGACAAGTGGATATCTTGGCTGAAGAAGGACATATCAGTAAGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAA
 CAACGGAAAGTAGTCCAACCTTTT

SEQ ID NO. 2311: SAG0163 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)
 CAGTAGAAGTAAATGCTCAAGATATTTATATCATTTCCCAAAGGTGATTGTTATGAATCTATATGCGTATTGATGATGAAAGCGGT
 TTATTGATGTTTTTGAGTTTAAATAGGATGGCTAGTCTTATTAGTCACTTTAAATTTGTGGCAGGCATGAACGTTGGAGAAAAAGAC
 GAAGTCAATTAGGTTCTTGTGACTATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCAAGTGTGGGAGATTATCGTGGTC
 AAGAATCTTTAGTTATTCGTACTTTGTATTTCAGGTCTATCAGGACTTAAAATATTTGGTTTGATAATATAAAGCAAATGAAGGAAGTAC
 TGTGTGCAAGAGGGCTATATCTTTTTCCGGCCCTGTGGGGAGTGGTAAACAACCTCTCATGTATCAATTAGCTTCAGAAGTATTTA
 AAAATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAACCTCAATTGAATGAGGATATTGGAA
 TGACTTATGATGCTTTAATCAAACGTCTTTACGGCATCGTCCAGATATTTTAAATTATCGGAGAGATTAGAGATCAAGCGACGGCCC
 GTGCTGTTATTCGTGCAAGTTTAAACGGGAGTAATGGTTTTTCTACTATTTCATGCTAAAAGTATTTCCCGGAGTCTATGATAGGCTTA
 TAGAATTAGGGGTTAACTATCAAGAGTTAGAAAATAGTCTAAAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTG
 ACTTTGAGACAAGTAACTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAAGGACATATCAGTA
 AGAAACAGGCACAAGTCGAAAAAATTATCCCTCAAGAAACAACGGAAAGTAGTCCAACCTTTT

**Table 23: Comparative Sequences relating to SAG01
(competence protein Cg1A)**

SEQ2301	-----GGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2302	-----
SEQ2303	TTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2304	-----GATATT
SEQ2305	TTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2306	TTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2307	-----
SEQ2308	-----TCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2309	TTCAATCATTAGCAAAGCAAGTCATTCATCAGGCAGTAGAAGTAAATGCTCAAGATATT
SEQ2310	-----
SEQ2311	-----CAGTAGAAGTAAATGCTCAAGATATT
SEQ2301	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2302	-----GGTGA-TTGTTATGAA-ACCTCTACTATTGCGTATTTGATGATGA
SEQ2303	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2304	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2305	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2306	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2307	-----AGGTGA-TTGTTATGAAATCTCTATA----TGCGTATT-GATGATGA
SEQ2308	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2309	ATATCATTCCCAAAGGTGA-TTGTTATGAA-CTCTATA----TGCGTATT-GATGATGA
SEQ2310	-----TGACTTGTATGAAACTCTATA----TGCGTATTTGATGATGA
SEQ2311	ATATCATTCCCAAAGGTGA-TTGTTATGAA-TTCTATA----TGCGTATT-GATGATGA
SEQ2301	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2302	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2303	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2304	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2305	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2306	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2307	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2308	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2309	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2310	AAAGGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2311	AA-GGCGGTTTATTGATGTTTTTTGAGTTTAATAGGATGGCTAGTCTTATTAGTCACTTTA
SEQ2301	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2302	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2303	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2304	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2305	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2306	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2307	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2308	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2309	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2310	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT
SEQ2311	AATTTGTGGCAGGCATGAACGTTGGAGAAAAAAGACGAAGTCAATTAGGTTCTTGTGACT

**Tabl 23: Comparative Sequences relating t SAG0163
(competence pr tein Cg1A)**

SEQ2301	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2302	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2303	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2304	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2305	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2306	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2307	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCAAGTGTGGGAGATTATCGTG
SEQ2308	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2309	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCGAGTGTGGGAGATTATCGTG
SEQ2310	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCAAGTGTGGGAGATTATCGTG
SEQ2311	ATGAAGTGTGAGAGGGAAGACTGGTTTCATTACGACTATCAAGTGTGGGAGATTATCGTG
SEQ2301	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2302	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2303	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2304	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2305	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2306	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2307	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2308	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2309	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2310	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2311	GTCAAGAATCTTTAGTTATTTCGTATTTTGTATTTCAGGTCATCAGGACTTAAATATTGGT
SEQ2301	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2302	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2303	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2304	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2305	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2306	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2307	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2308	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2309	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2310	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2311	TTGATAATATAAAGCAAATGAAGGAAGTACTGGGTACAAGAGGGCTATATCTTTTTTCCG
SEQ2301	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2302	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2303	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2304	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2305	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2306	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2307	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2308	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2309	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2310	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA
SEQ2311	GCCCTGTGGGGAGTGGTAAACAACTCTCATGTATCAATTAGCTTCAGAAGTATTTAAAA

**Table 23: Comparative Sequences relating to SAG0163
(competenc protein Cg1A)**

SEQ2301	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2302	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2303	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2304	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2305	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2306	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2307	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2308	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2309	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2310	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2311	ATAAGCAAATTATCACGATTGAAGATCCGGTAGAAATCAAGAATGACAAGATGTTACAAC
SEQ2301	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2302	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2303	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2304	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2305	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2306	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2307	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2308	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2309	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2310	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2311	TCCAATTGAATGAGGATATTGGAATGACTTATGATGCTTTAATCAAACGTCTTTACGGC
SEQ2301	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2302	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCTCGTGCTGTT
SEQ2303	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2304	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2305	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2306	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2307	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2308	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2309	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2310	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2311	ATCGTCCAGATATTTTAATTATCGGAGAGAT-TAGAGATCAAGCGACGGCCCGTGCTGTT
SEQ2301	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2302	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2303	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2304	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2305	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2306	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2307	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2308	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2309	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2310	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2311	ATTCGTGCAAGTTTAACGGGAGTGATGGTTTTTTCTACTATTATCATGCTAAAAGTATTCCC
SEQ2301	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2302	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2303	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2304	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2305	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2306	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2307	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2308	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2309	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2310	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2311	GGAGTCTATGATAGGCTTATAGAATTAGGGGTAACTATCAAGAGTTAGAAAATAGTCTA
SEQ2301	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2302	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2303	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2304	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2305	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2306	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2307	AAATTAATAGCATATCAACGTTTAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT

**Table 23: Comparative Sequences relating t SAG0163
(competence protein Cg1A)**

SEQ2308	AAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2309	AAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAGGT
SEQ2310	AAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAAGT
SEQ2311	AAATTAATAGCATATCAACGTTTAAATTGGAGGAGGAAGCCTAATTGACTTTGAGACAAGT
SEQ2301	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2302	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2303	AATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2304	AATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2305	AATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2306	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2307	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2308	AATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2309	AATTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2310	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2311	AACCTTTAAAAAACACTCATCAGACAAGTGGAAATAGACAAGTGGATATCTTGGCTGAAGAA
SEQ2301	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2302	GGATATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2303	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2304	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2305	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2306	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2307	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2308	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2309	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2310	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2311	GGACATATCAGTAAGAAACAGGCACAAGT-CGAAAAAATTATCCCTCAAGAAACAACGGA
SEQ2301	AAGTAGTCCAACCTTTT-----
SEQ2302	AAGTAGTCCAACCTTTT-----
SEQ2303	AAGTAGTCCAACCTTTT-----
SEQ2304	AAGTAGTCCAACCTTTT-----
SEQ2305	AAGTAGTCCAACCTTTT-----
SEQ2306	AAGTAGTCCAACCTTTT-----
SEQ2307	AAGTAGTCCAACCTTTT-----
SEQ2308	AAGTAGTCCAACCTTTT-----
SEQ2309	AAGTAGTCCAACCTTTT-----
SEQ2310	AAGTAGTCCAACCTTTT-----
SEQ2311	AAGTAGTCCAACCTTTT-----
>SEQ ID NO 2350:63_090 frame: 2	
AVEVNAQDIYIIPKGDCELYMRIDDERFIDVFEFNRMASLISHFKFVAGMNVGEKRRS	
QLGSCDYELSEGRVLSRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIKQMKVGLGTR	
GLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQNLNEDIGMTYDAL	
IKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVYDRLIELGVNYQ	
ELENKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHISKKQAQVEKII	
PQETTESPTF	
>SEQ ID NO 2351:63_1169NT frame: 3	
.LL.NLYYCVFDDERRFIDVFEFNRMASLISHFKFVAGMNVGEKRRS	
QLGSCDYELSEGRVLSRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIKQMKVGLGTR	
GLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQNLNEDIGMTYDAL	
IKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVYDRLIELGVNYQ	
ELENKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGYISKQAQVEKIIPQETTESPTF	
>SEQ ID NO 2352:63_18RS21 frame: 1	
VQSLAQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERFIDVFEFNRMASLISHFKFV	
AGMNVGEKRRS	
QLGSCDYELSEGRVLSRLSSVGDYRGQESLVIRILYSGHQDLKYWFDN	
IKQMKVGLGIRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQ	
NEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY	
DRLIELGVNYQELENKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI	
SKKQAQVEKIIPQETTESPTF	
>SEQ ID NO 2353: 63_2603 frame: 1	
DIYIIPKGDCELYMRIDDERFIDVFEFNRMASLISHFKFVAGMNVGEKRRS	
QLGSCDYELSEGRVLSRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIKQMKVGLGIRGLYLFSG	

**Table 23: Comparative Sequences relating to SAG0165
(competence protein Cg1A)**

PVSGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQLNEDIGMTYDALIKLSLRH
RPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVYDRILIELGVNYQELENLSK
LIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHISKQAQVRKNYPSPRNGK
.SNF

>SEQ ID NO 2354:63_A909 frame: 1

VQSLAKQVIHQAVEVNAQDIYIIPKGDCYELMRIDDERRFIDVFEFNRMASLISHFKFV
AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDN
IKQMKVGLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
NEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
DRILIELGVNYQELENLSKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SKQAQVEKIIPQETTESPTF

>SEQ ID NO 2355:63_CJB110 frame: 1

VQSLAKQVIHQAVEVNAQDIYIIPKGDCYELMRIDDERRFIDVFEFNRMASLISHFKFV
AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDN
IKQMKVGLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
NEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
DRILIELGVNYQELENLSKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SKQAQVEKIIPQETTESPTF

>SEQ ID NO 2356:63_CJB110 frame: 1

VQSLAKQVIHQAVEVNAQDIYIIPKGDCYELMRIDDERRFIDVFEFNRMASLISHFKFV
AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDN
IKQMKVGLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
NEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
DRILIELGVNYQELENLSKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SKQAQVEKIIPQETTESPTF

>SEQ ID NO 2357: 63_H36B frame: 1

SLAKQVIHQAVEVNAQDIYIIPKGDCYELMRIDDERRFIDVFEFNRMASLISHFKFVAG
MNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIK
QMKVGLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQLNE
DIGMTYDALIKLSLRHRPDILIIGE

>SEQ ID NO 2358:63_JM9130013 frame: 1

VQSLAKQVIHQAVEVNAQDIYIIPKGDCYELMRIDDERRFIDVFEFNRMASLISHFKFV
AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDN
IKQMKVGLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
NEDIGMTYDALIKLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
DRILIELGVNYQELENLSKLIAYQRLIGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SKQAQVEKIIPQETTESPTF

>SEQ ID NO 2359:63_M732 frame: 3

TCYETLYAYLMMKRRFIDVFEFNRMASLISHFKFVAGMNVGEKRRSQLGSCDYELSEGRL
VSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIK.MKEVLCARGLYLFSGPVGSGKT
TLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQLNEDIGMTYDALIKLSLRHRPDILII
GEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVYDRILIELGVNYQELENLSKLIAYQRL
IGGSLIDFETS NFKKHSSDKWNRQVDILAEEGHISKQAQVEKIIPQETTESPTF

>SEQ ID NO 2360:63_M781 frame: 3

VEVNAQDIYIIPKGDCYEFYMRIDDERRFIDVFEFNRMASLISHFKFVAGMNVGEKRRSQ
LGSCDYELSEGRLVSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIKQMKVLCARG
LYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQLNEDIGMTYDALI
KLSLRHRPDILIIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVYDRILIELGVNYQE
LENSKLIAYQRLIGGSLIDFETS NFKKHSSDKWNRQVDILAEEGHISKQAQVEKIIP
QETTESPTF

>SEQ ID NO 2361:63_COH1 frame: 3

VIVMKFYMRIDDERRFIDVFEFNRMASLISHFKFVAGMNVGEKRRSQLGSCDYELSEGRL
VSLRLSSVGDYRGQESLVIRILYSGHQDLKYWFDNIK

Table 23: Comparative Sequences relating to SAG0105
(competence protein Cg1A)

SEQ2350	-----AVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2351	-----LLNLYYCVFDDERRRFIDVFEFNRMASLISHFKFV
SEQ2352	QSLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2353	-----DIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2354	QSLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2355	QSLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2356	QSLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2357	-SLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2358	QSLAKQVIHQAVEVNAQDIYIIPKGDCELYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2359	-----TCYETLYAYLMMKRRRFIDVFEFNRMASLISHFKFV
SEQ2360	-----VEVNAQDIYIIPKGDCELYFMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2361	-----VIVMKFYMRIDDERRFIDVFEFNRMASLISHFKFV
SEQ2350	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2351	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2352	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2353	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2354	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2355	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2356	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2357	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2358	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2359	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2360	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2361	AGMNVGEKRRSQLGSCDYELSEGRLVSLRLSSVGDYRGQESLIRILYSGHQDLKYWFDN
SEQ2350	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2351	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2352	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2353	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2354	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2355	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2356	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2357	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2358	IKQMKEVLGTRGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2359	IK-MKEVLCARGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2360	IKQMKEVLCARGLYLFSGPVGSGKTTLMYQLASEVFKNKQIITIEDPVEIKNDKMLQLQL
SEQ2361	IK-----
SEQ2350	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
SEQ2351	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2352	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2353	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2354	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2355	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
SEQ2356	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
SEQ2357	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSISGVY
SEQ2358	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2359	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2360	EDIGMTYDALIKLSLRHRPDILIGEIRDQATARAVIRASLTGVMVFSTIHAKSIPGVY
SEQ2361	-----
SEQ2350	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2351	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETSNFKKHSSDKWNRQVDILAEEGYI
SEQ2352	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2353	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2354	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2355	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2356	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2357	-----
SEQ2358	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETGNFKKHSSDKWNRQVDILAEEGHI
SEQ2359	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETSNFKKHSSDKWNRQVDILAEEGHI
SEQ2360	RLIELGVNYQELENLKLIAAYQRLIGGGSLIDFETSNFKKHSSDKWNRQVDILAEEGHI
SEQ2361	-----

**Table 23: C mparative Sequences relating t SAG0163
(competence protein Cg1A)**

SEQ2350	KKQAQVEKIIPQETTESPTF
SEQ2351	KKQAQVEKIIPQETTESPTF
SEQ2352	KKQAQVEKIIPQETTESPTF
SEQ2353	KKQAQVRKNYPSRNGKSNF-
SEQ2354	KKQAQVEKIIPQETTESPTF
SEQ2355	KKQAQVEKIIPQETTESPTF
SEQ2356	KKQAQVEKIIPQETTESPTF
SEQ2357	-----
SEQ2358	KKQAQVEKIIPQETTESPTF
SEQ2359	KKQAQVEKIIPQETTESPTF
SEQ2360	KKQAQVEKIIPQETTESPTF
SEQ2361	-----

**Figure 24: C mparative Sequences relating to SAG0290
(ABC transporter, substrate-binding protein)**

SEQ ID NO. 2401: SAG0290 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGACCCTATATCCCGTTCAAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACAGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAAGATCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAG
TAAACAATATTTCCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ ID NO. 2402: SAG0290 FROM THE 18RS21 GBS TYPE II STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TRAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAAATGGTACTTTGGCACGTTTAAG
TAAACAATATTTCCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ ID NO. 2403: SAG0290 FROM THE 2603 V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)

ATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGA
CAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCAGCTAATGATTTTTCA
TACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAAATTATGCCGTAGTAGG
GAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAGTTTTATCTGGCGTTA
ACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATAAAAAATCAAATATGTT
TCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGACTTTATCCTATATGA
TGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTCCTTTGAAAGGTAAAA
TTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAGATAAAAAAG

SEQ ID NO. 2404: SAG0290 FROM THE 090 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)
GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTCCT
TTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTTA
CAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAAATGGTACTTTGGCACGTTTAAGTAACAATATTTCCGGT
GGAGATTACGTTTCAAACATTGATAAA

0114144-012702

**Figure 24: Comparative Sequences relating to SAG0290
(ABC transporter, substrate-binding protein)**

SEQ ID NO. 2405: SAG0290 FROM THE A909 GBS TYPE Ia STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATNNTAATAAAAAACCANTA
AAAATNAAATATGTTTCTGGGACAACCTGGTGTTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGT

SEQ ID NO. 2406: SAG0290 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAAATGGTACTTTGGCACGTTTAAG
TAAACAATATTTTCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ ID NO. 2407: SAG0290 FROM THE COH1 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATATAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACAGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAAGATCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAG
TAAACAATATTTTCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ ID NO. 2408: SAG0290 FROM THE H36b GBS TYPE Ib STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAAATGGTACTTTGGCACGTTTAAG
TAAACAATATTTTCGGTGGAGATTACGTTTCAAACATTGATAAA

**Figure 24: Comparative Sequences relating to SAG0290
(ABC transporter, substrate-binding protein)**

SEQ ID NO. 2409: SAG0290 FROM THE JM9130013 GBS STRAIN VIII (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATACAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACCGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAAGACCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTAATAAAGTTTTGAAAGAAAATGGTA

SEQ ID NO. 2410: SAG0290 FROM THE M732 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATATAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACAGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAAGATCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAG
TAAACAATATTTTCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ ID NO. 2411: SAG0290 FROM THE M781 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GTATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCACCATTACTTA
TCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCTGTTTTTAAAGGTAGTAAGTACA
AAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCAACAGGTATTGATGCAGGGAAATTTGATTTATCA
GCTAATGATTTTTTCATATAATAAAGAAAGAGCAGAAAAATATCTCTTCTCAGATCCTATATCCCGTTCAA
TTATGCCGTAGTAGGGAAGAAGGGGAGCCATTACAAATCATTAAAGTGACCTCTCTGGAAAATCAACAGAAG
TTTTATCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAACCAATA
AAAATCAAATATGTTTCTGGGACAACCTGGTGTACTAGCAGATTAAAAAATATTGAGAGTGGGAAAATTGA
CTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAAGATCAATCATTAACTTAAGCGTTTCTC
CTTTGAAAGGTAAAATTGGTAATAATAAGGATGGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGT
AAAACCTCTACAGAAATTTATAAATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAG
TAAACAATATTTTCGGTGGAGATTACGTTTCAAACATTGATAAA

SEQ2401	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2402	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2403	-----
SEQ2404	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2405	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2406	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2407	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2408	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2409	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2410	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA
SEQ2411	TATCAGTTCAGGCGTCAGAGAAAGTAGAACTTAAAGTAGCTACAGATTCTGACACGGCA

SEQ2401	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2402	CATTACTTATRAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2403	-----ATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2404	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2405	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2406	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2407	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCT
SEQ2408	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT
SEQ2409	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGATGTTGATGTTGTCAAAGCT

**Figure 24: Comparative Sequences relating to SAG02
(ABC transporter, substrate-binding protein)**

SEQ2410	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCT
SEQ2411	CATTACTTATCAAAAAGACGGGAAATTCAAAGGTTATGACGTTGATGTTGTCAAAGCT
SEQ2401	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2402	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2403	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2404	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2405	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2406	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2407	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2408	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2409	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2410	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2411	GTTTTTAAAGGTAGTAAGTACAAAGTAACCTTCAAGACAGTTCCTTTTGATACTATTTCA
SEQ2401	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2402	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2403	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2404	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2405	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2406	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2407	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2408	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2409	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2410	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2411	ACAGGTATTGATGCAGGGAAATTTGATTATCAGCTAATGATTTTTCATACAATAAAGAA
SEQ2401	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2402	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2403	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2404	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2405	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2406	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2407	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2408	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2409	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2410	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2411	AGAGCAGAAAAATATCTCTTCTCAGACCTATATCCCGTTCAAATTATGCCGTAGTAGGG
SEQ2401	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2402	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2403	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2404	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2405	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2406	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2407	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2408	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2409	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2410	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2411	AAGAAGGGGAGCCATTACAAATCATTAAGTGACCTCTCTGGAAAATCAACAGAAGTTTTA
SEQ2401	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2402	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2403	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2404	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2405	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2406	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2407	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2408	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2409	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2410	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2411	TCTGGCGTTAACTATGCACAGGTTCTAGAAAATTGGAATAAAAATCATCCTAATAAAAAA
SEQ2401	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2402	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2403	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2404	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2405	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2406	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2407	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2408	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT
SEQ2409	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAATATT

(ABC transporter, substrate-binding protein)

SEQ2410	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAAATATT
SEQ2411	CCAATAAAAAATCAAATATGTTTCTGGGACAACTGGTGTACTAGCAGATTAAAAAATATT
SEQ2401	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAA
SEQ2402	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2403	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2404	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2405	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2406	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2407	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAA
SEQ2408	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2409	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCCGACTATATTGTAAAA
SEQ2410	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAA
SEQ2411	GAGAGTGGGAAAATTGACTTTATCCTATATGATGCCATTTTCATCTGACTATATTGTAAAA
SEQ2401	GATCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2402	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2403	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2404	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2405	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2406	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2407	GATCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2408	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2409	GACCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2410	GATCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2411	GATCAATCATTAAGCTTAAGCGTTTCTCCTTTGAAAGGTAAAATTGGTAATAATAAGGAT
SEQ2401	GGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2402	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2403	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2404	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2405	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2406	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2407	GGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2408	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2409	GGACTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2410	GGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2411	GGATTAGAATACCTCCTTTTACCAAAGATAAAAAAGGTAAAACCTCTACAGAAATTTATA
SEQ2401	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2402	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2403	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2404	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2405	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2406	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2407	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2408	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2409	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2410	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2411	ATAAGCGTATTAAAGTTTTGAAAGAAGATGGTACTTTGGCACGTTTAAAGTAAACAATAT
SEQ2401	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2402	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2403	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2404	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2405	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2406	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2407	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2408	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2409	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2410	TCGGTGGAGATTACGTTTCAAACATTGATAAA-----
SEQ2411	TCGGTGGAGATTACGTTTCAAACATTGATAAAAGTRCMARATVSTNCSRATNGTSAGABC

**Figure 24: Comparative Sequences relating to SAG02
(ABC transporter, substrate-binding protein)**

```
SEQ2401 -----
SEQ2402 -----
SEQ2403 -----
SEQ2404 -----
SEQ2405 -----
SEQ2406 -----
SEQ2407 -----
SEQ2408 -----
SEQ2409 -----
SEQ2410 -----
SEQ2411 RANSRTRSTBSTRATBNDNGRTN
```

>SEQ ID NO 2450: 8_1169NT frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKEDGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2451:8_18RS21 frame: 1
VSVQASEKVELKVATDSDTAPFTYXKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKENGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2452:8_2603 frame: 2
FKGYDVDVVKAVFKGSKYKVTFTKVPFDTISTGIDAGKFDLSANDFSYNKERAEEKYLFSD
PISRSNYAVVGKKGSHYKSLSDLGKSTEVLSGVNYAQVLENWNNKHPNKKPIKIKYVSG
TTGVTSLRKNIESGKIDFILYDAISSDYIVKDQSLNLSVSPLKKGKIGNNKDGLLEYLLLP
DKK

>SEQ ID NO 2453:8_090 frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKENGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2454:8_A909 frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKR

>SEQ ID NO 2455: 8_CJB110 frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKENGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2456: 8_COH1 frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKEDGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2457:8_H36B frame: 1
VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFTKVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRKNIESGKIDFILYDAISSDYIVK
DQSLNLSVSPLKKGKIGNNKDGLLEYLLLPKDKKGKTLQKFINKRIKVLKENGTLARLSKQY
FGGDYVSNIDK

Figure 24: Comparative Sequences relating to SAG02
(ABC transporter, substrate-binding protein)

>SEQ ID NO 2458:8_JM9130013 frame: 1

VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRNKVLKENG

>SEQ ID NO 2459:8_M732 frame: 1

VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY
FGGDYVSNIDK

>SEQ ID NO 2460:8_M781 frame: 1

VSVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY
FGGDYVSNIDK

SEQ2450	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2451	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2452	-----FKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2453	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2454	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2455	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2456	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2457	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2458	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2459	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2460	SVQASEKVELKVATDSDTAPFTYQKDGKFKGYDVDVVKAVFKGSKYKVTFKTVPFDTIS
SEQ2450	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2451	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2452	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2453	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2454	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2455	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2456	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2457	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2458	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2459	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2460	TGIDAGKFDLSANDFSYNKERAEEKYLFSDPISRSNYAVVGKKGSHYKSLSDLGKSTEV
SEQ2450	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2451	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2452	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2453	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2454	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2455	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2456	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2457	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2458	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2459	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2460	SGVNYAQVLENWNNKHPNKKPIKIKYVSGTTGVTSLRNKIESGKIDFILYDAISSDYIVK
SEQ2450	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY
SEQ2451	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKENGTLARLSKQY
SEQ2452	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKK-----
SEQ2453	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKENGTLARLSKQY
SEQ2454	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKR-----
SEQ2455	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKENGTLARLSKQY
SEQ2456	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY
SEQ2457	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKENGTLARLSKQY
SEQ2458	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRNKVLKENG-----
SEQ2459	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY
SEQ2460	DQSLNLSVSPKKGKIGNNKDGLLEYLLLPKDKKGGKTLOKFINKRIKVLKEDGTLARLSKQY

**Figure 24: Comparative Sequences relating to SAG02
(ABC transporter, substrate-binding protein)**

SEQ2450	GGDYVSNIDK
SEQ2451	GGDYVSNIDK
SEQ2452	-----
SEQ2453	GGDYVSNIDK
SEQ2454	-----
SEQ2455	GGDYVSNIDK
SEQ2456	GGDYVSNIDK
SEQ2457	GGDYVSNIDK
SEQ2458	-----
SEQ2459	GGDYVSNIDK
SEQ2460	GGDYVSNIDK

3

S

T

5

13

2

2

**Table 25: Comparative Sequences relating to SAG0368
(protein of unknown function)**

SEQ ID NO. 2504: SAG0368 FROM THE 2603 V/R GBS TYPE V STRAIN

TATAATTTTTCGACTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCTATTGAA
GAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAAT
AGCGATTCTATGATCTTAGTCACTATAAATCCTAAAACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTG
ATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGAGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGT
GCGGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGA
TTAGTTGATTTAGTCAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAAT
GAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTATTCTCGTATG
CGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAGACAACGTAAGTAATTCAAAAGTCCTTAAAAAATA
TTGGCGTTAAATAGTATTAGTTTCATACAAAAAATTCTTTCCGCAGTAAGTAATAACATGCAAACTAATATTGAGATA
TCATCAAAAACGATTCCCTAATTTGTTAGCTTATAAAGATTTCATTGGAACATATTAATCTTATCAGTTGAAGGGTGAA
GACGCTACTTTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAG
AAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGGATTCTATATGAAGATTACTATGGTACTACTGCT
AGTAATGATTCTTCTACTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGT
TACAGTGGTAATACTACTTATAGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTCCTGCT
AGTAATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCAATAATCATAACGGGGCTGCAACG
CCTAATCCA

SEQ ID NO. 2505: SAG0368 FROM THE A909 GBS TYPE Ia STRAIN

TATAATTTTTCGACTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCTATTGAA
GAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAAT
AGCGATTCTATGATCTTAGTCACTATAAATCCTAAAACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTG
ATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGAGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGT
GCGGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGA
TTAGTTGATTTAGTCAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAAT
GAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTATTCTCGTATG
CGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAGACAACGTAAGTAATTCAAAAGTCCTTAAAAAATA
TTGGCGTTAAATAGTATTAGTTTCATACAAAAAATTCTTTCCGCAGTAAGTAATAACATGCAAACTAATATTGAGATA
TCATCAAAAACGATTCCCTAATTTGTTAGCTTATAAAGATTTCATTGGAACATATTAATCTTATCAGTTGAAGGGTGAA
GACGCTACTTTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAG
AAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGGATTCTATATGAAGATTACTATGGTACTACTGCT
AGTAATGATTCTTCTACTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGT
TACAGTGGTAATACTACTTATAGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTCCTGCT
AGTAATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCAATAATCATAACGGGGCTGCAACG
CCTAATCCA

**SEQ ID NO. 2506: SAG0368 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE
COMPLEMENT)**

TATAATTTTTCGACTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCTATTGAA
GAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAAT
AGCGATTCTATGATCTTAGTCACTATAAATCCTAAAACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTG
ATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGAGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGT
GCGGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGA
TTAGTTGATTTAGTCAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAAT
GAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTATTCTCGTATG
CGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAGACAACGTAAGTAATTCAAAAGTCCTTAAAAAATA
TTGGCGTTAAATAGTATTAGTTTCATACAAAAAATTCTTTCCGCAGTAAGTAATAACATGCAAACTAATATTGAGATA
TCATCAAAAACGATTCCCTAATTTGTTAGCTTATAAAGATTTCATTGGAACATATTAATCTTATCAGTTGAAGGGTGAA
GACGCTACTTTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAG
AAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGGATTCTATATGAAGATTACTATGGTACTACTGCT
AGTAATGATTCTTCTACTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGT
TACAGTGGTAATACTACTTATAGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTCCTGCT
TAGTAATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCAATAATCATAACGGGGCTGCAAC
GCCTAATCCA

**Table 25: Comparative Sequences relating t SAG0368
(protein of unknown function)**

SEQ ID NO. 2507: SAG0368 FROM THE COH1 GBS TYPE III STRAIN (REVERSE COMPLEMENT)

GATTTTAAGCTAGATAAATCAAAAAGTCATGCTATTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGTGTGGAC
ACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAATCCTAAACT
AATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGC
GTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGTGCAGAAATGGCATTGATGACTGTTCAAGACTTATTAGAT
ATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGATTAGTTGATTTGGTCAATGCTGTTGGTGGTATAACAGTA
ACTAATAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACAT
AAAAATAATGGAGAACAAGCACTTGTTTTATTCTCGTATGCGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAA
AGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCTT
TCCGCGTAGTAATAACATGCAAACTAATATTGAGATATCATCAAAAACGATTCTAATTTGTTAGCTTATAAAGAT
TCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTACTCTATCAGATGGTGGCTCTTATCAAATTTTA
ACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAGAAAGAGCTGGATAAAAAGCGTAGTAAACTCTGAAGACA
AGCGCGATTCTATATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACAAGAGAAT
TATTATTATACAACACCCCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATACTACTTATAGTTCTGAGACTAATCA
AACAACCTCATCAAAGTTACTATAATAGTAGCACTCCTGCTAGTAACATATAGCAGTAACACTAACACAGGCTCAGGCTGA
TTCAAGTGAAGTGTTAATAATTATAACGGGGCTGCAACGCCAATCCAAACACAGGAACGCAACCAGTACCAGGTCA
AACTAATCCA

SEQ ID NO. 2508: SAG0368 FROM THE H36b GBS TYPE Ib STRAIN

TATAATTTTTTCGACTAATGAATTGTCTAAGACTTTTAAAGATTTTAAGCTAGCTAAATCAAAAAGTCATGCTATTGAA
GAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAAT
AGCGATTCTATGATCTTAGTCACTATAAATCCTAAAACATAAAACAACGATGACAAGCTTAGAACGTGACGTATTG
ATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGAGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGT
GCGGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGA
TTAGTTGATTTAGTCAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAAT
GAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTTATTCTCGTATG
CGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATA
TTGGCGTTAAATAGTA

SEQ ID NO. 2509: SAG0368 FROM THE _____

TTAGTTCATACAAAAAATCTTTCCGCGTAGTAATAACATGCAAACTAATATTGAGATATCATCAAAAACGATTCT
CTAATTTGTTAGCTTATAAAGATTCAATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTACTTTATCAG
ATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAAATAGAATTAAGAAAGAACTGGATAAAA
AGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTA
CTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATACTA
CTTATAGTTCTGAGACTAATCAAACAACCTCATCAAAATTACTATAATAGTAGCACTCCTGCTAGTAACATATAGCAGTA
ACACTAACACAGGTGAGGCTGATTCAAGTGGAAAGTGCAATAATCATAACGGGGCTGCAACGCCAATCCA

SEQ ID NO. 2510: SAG0368 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)

TATAATTTTTTCGACTAATGAATTGTCTAAGACTTTTAAAGATTTTAAGCTAGCTAAATCAAAAAGTCATGCTATTGAA
GAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAAT
AGCGATTCTATGATCTTAGTCACTATAAATCCTAAAACATAAAACAACGATGACAAGCTTAGAACGTGACGTATTG
ATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGAGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGT
GCGGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGA
TTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAAT
GAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTTATTCTCGTATG
CGCTATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATA
TTGGCGTTAAATAGTATTAGTTCATACAAAAAATCTTTCCGCGTAGTAATAAATCCTTATCAGTTGAAGGGTGAA
TCATCAAAAACGATTCTAATTTGTTAGCTTATAAAGATTCAATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAA
GACGCTACTTTTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAAATAGAATTAAG
AAAGAACCTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACTACTGCT
AGTAATGATTCTTCTACTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGT
TACAGTGGTAATACTACTTATAGTTCTGAGACTAATCAAACAACCTCATCAAAATTACTATAATAGTAGCACTCCTGCT
AGTAACATAGCAGTAACACTAACACAGGTGAGGCTGATTCAAGTGGAAAGTGCAATAATCATAACGGGGCTGCAACG
CCTAATCCA

Table 25: Comparative Sequences relating to SAG0368
(protein of unknown function)

SEQ ID NO. 2511: SAG0368 FROM THE M781 GBS TYPE III STRAIN (REVERSE
COMPLEMENT)

TTCAATACTATTAATGGGTGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAATAGCGATTCTATGAT
CTTAGTCACTATAAATCCTAAACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGG
TCCCAAAAATAATGGACAGACTGGCGTAGAAGCAAAGCTAAATGCAGCCTATGCTTCTGGTGGTGCAGGAAATGGCATT
GATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGATTAGTTGATTGGT
CAATGCTGTTGGTGGTATAACAGTAACATAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAA
GGCTGTTGTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCGCTATGATGATCC
AGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATATTGGCGTTAAATAG
TATTAGTTCATACAAAAAATCTTTCCGCAGTAAGTAATAACATGCAAATAATATTGAGATATCATCAAAAACGAT
TCCTAATTTGTTAGCTTATAAAGATTCATTGGAACATATTAATCTTATCAGTTGAAGGGTGAAGACGCTACTCTATC
AGATGGTGGCTCTTATCAAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAGAAAGAGCTGGATAA
AAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTC
TACTTATTCATCAACACAAGAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
TACTTATAGTTCTGAGACTAATCAAACAACCTCATCAAAGTTACTATAATAGTAGCACTCCTGCTAGTAACATAGCAG
TAACACTAACACAGGTCAGGCTGATTCAAGTGGAAAGTGTTAATAATTATAACGGGGCTGCAACGCCTAATCCAAACAC
AGGAACGCAACCAGTACCAGGTCAAATAATCCA

SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	ATTTTAAGCTAGATAAATCAAAAAGTCATGCTATTGAAGAAACAAAGCCGTTTTCAATA
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	-----TTCAATA

SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	TATTAATGGGTGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAATAGC
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	TATTAATGGGTGTGGACACAGGTTTCAGAGCATCGAAAATCTAAGTGGTCAGGAAATAGC

SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	ATTCTATGATCTTAGTCACTATAAATCCTAAACTAATAAAACAACGATGACAAGCTTA
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	ATTCTATGATCTTAGTCACTATAAATCCTAAACTAATAAAACAACGATGACAAGCTTA

**Table 25: Comparative Sequences relating to SAG0306
(protein of unknown function)**

SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	AACGTGACGTATTGATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGCGTAGAA
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	AACGTGACGTATTGATTAAATTGAGTGGTCCCAAAAATAATGGACAGACTGGCGTAGAA
SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	CAAAGCTAAATGCAGCCTATGCTTCTGGTGGTGCGGAAATGGCATTGATGACTGTTCAA
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	CAAAGCTAAATGCAGCCTATGCTTCTGGTGGTGCGGAAATGGCATTGATGACTGTTCAA
SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	ACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGATTAGTTGAT
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	ACTTATTAGATATTAATGTTGATTACTTTATGCAAATTAATATGCAAGGATTAGTTGAT
SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	TGGTCAATGCTGTTGGTGGTATAACAGTAACATAAAATTTGACTTTCCAATATCAATT
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	TGGTCAATGCTGTTGGTGGTATAACAGTAACATAAAATTTGACTTTCCAATATCAATT
SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	CTGCCAATGAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGA
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	CTGCCAATGAACCAGAGTACAAGGCTGTTGTTGAACCAGGGACACATAAAATAAATGGA

**Table 25: Comparative Sequences relating to SAG0300
(protein of unknown function)**

SEQ2501	-----
SEQ2502	-----
SEQ2503	-----
SEQ2504	-----
SEQ2505	-----
SEQ2506	-----
SEQ2507	AACAAGCACTTGTTTATTCTCGTATGCGCTATGATGATCCAGAGGGAGATTATGGGCGT
SEQ2508	-----
SEQ2509	-----
SEQ2510	-----
SEQ2511	AACAAGCACTTGTTTATTCTCGTATGCGCTATGATGATCCAGAGGGAGATTATGGGCGT
SEQ2501	-----TATAATTTTTTCG
SEQ2502	-----TATAATTTTTTCG
SEQ2503	-----TATAATTTTTTCG
SEQ2504	-----TATAATTTTTTCG
SEQ2505	-----TATAATTTTTTCG
SEQ2506	-----TATAATTTTTTCG
SEQ2507	AAAAAAGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATATTGGCGTTAAATAGT
SEQ2508	-----TATAATTTTTTCG
SEQ2509	-----
SEQ2510	-----TATAATTTTTTCG
SEQ2511	AAAAAAGACAACGTGAAGTAATTCAAAAAGTCCTTAAAAAATATTGGCGTTAAATAGT
SEQ2501	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2502	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2503	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2504	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2505	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2506	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2507	TTAGTTCAT-ACAAAAAATTCTTTCCGCAGTAAGTAA--TAACATGCAAACATAATATT
SEQ2508	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2509	TTAGTTCAT-ACAAAAAATTCTTTCCGCAGTAAGTAA--TAACATGCAAACATAATATT
SEQ2510	CTAATGAATTGTCTAAGACTTTTAAAGATTTTAAAGCTAGCTAAATCAAAAAGTCATGCT
SEQ2511	TTAGTTCAT-ACAAAAAATTCTTTCCGCAGTAAGTAA--TAACATGCAAACATAATATT
SEQ2501	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2502	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2503	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2504	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2505	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2506	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2507	AGATATCATCAAAAACGATTCTTAATTTGTTAGCTTATAAAGATTCA---TTGGAACAT
SEQ2508	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2509	AGATATCATCAAAAACGATTCTTAATTTGTTAGCTTATAAAGATTCA---TTGGAACAT
SEQ2510	TTGAAGAAACAAAGCCGTTTTCAATACTATTAATGGGGGTGGACACAGGTTTCAGAGCAT
SEQ2511	AGATATCATCAAAAACGATTCTTAATTTGTTAGCTTATAAAGATTCA---TTGGAACAT
SEQ2501	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2502	GAAAATCTAAGTGGTTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2503	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2504	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2505	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2506	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2507	TTAAATCTTATC-AGTTGAAGGGTGAAGACGCTACTCTATCAG--ATGGTGGCTCTTAT
SEQ2508	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2509	TTAAATCTTATC-AGTTGAAGGGTGAAGACGCTACTTTATCAG--ATGGTGGCTCTTAT
SEQ2510	GAAAATCTAAGT-GGTCAGGAAATAGCGATTCTATGATCTTAGTCACTATAAAATCCTAA
SEQ2511	TTAAATCTTATC-AGTTGAAGGGTGAAGACGCTACTCTATCAG--ATGGTGGCTCTTAT
SEQ2501	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC

**Table 25: Comparative Sequences relating to SAG0368
(protein of unknown function)**

SEQ2502	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2503	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2504	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2505	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2506	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2507	AAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAGAAAGAGCTGGAT
SEQ2508	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2509	AAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAGAAAGAGCTGGAT
SEQ2510	ACTAATAAAACAACGATGACAAGCTTAGAACGTGACGTATTGATTAAATTGAGTGGTCC
SEQ2511	AAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAATAGAATTAAGAAAGAGCTGGAT
SEQ2501	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2502	AAAAATAATGGACAGACTGGCGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2503	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2504	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2505	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2506	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2507	AAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACT
SEQ2508	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2509	AAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACT
SEQ2510	AAAAATAATGGACAGACTGGAGTAGAAGCAAAG--CTAAATGCAGCCTATGCTTCTGGT
SEQ2511	AAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCTATATGAAGATTACTATGGTACT
SEQ2501	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2502	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2503	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2504	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2505	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2506	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2507	CTGCTAGTAATGATTCTTCTACTTATTCATCAAC-ACAAGAGAATTATTATTAT-ACAA
SEQ2508	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2509	CTGCTAGTAATGATTCTTCTACTTATTCATCAAC-ACAAGAGAATAATTATAAT-ACAA
SEQ2510	GTGC-GGAAATGGCATTGATGACTGTTCAAGACTTATTAGATATTAATGTTGATTACTT
SEQ2511	CTGCTAGTAATGATTCTTCTACTTATTCATCAAC-ACAAGAGAATAATTATAAT-ACAA
SEQ2501	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2502	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2503	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2504	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2505	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2506	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2507	ACCCTTATTCAGAAGCACCACCAAGTTACAGTGGT-AATACTACTTATAGTT---CTGA
SEQ2508	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2509	ACC-TTATTCAGAAGCACCACCAAGTTACAGTGGT-AATACTACTTATAGTT---CTGA
SEQ2510	ATGCAAATTAATATGCAAGGATTAGTTGATTAGTCAATGCTGTTGGTGGTATAACAGT
SEQ2511	ACC-TTATTCAGAAGCACCACCAAGTTACAGTGGT-AATACTACTTATAGTT---CTGA
SEQ2501	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2502	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2503	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2504	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2505	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2506	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2507	ACTAATCAAAC-AACTCATCAA----AGTTACTAT-AATAG--TAGCACTCCTGCTAGT
SEQ2508	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2509	ACTAATCAAAC-AACTCATCAA----AATTACTAT-AATAG--TAGCACTCCTGCTAGT
SEQ2510	ACTAATAAAATTTGACTTTCCAATATCAATTGCTGCCAATGAACCAGAGTACAAGGCTGT
SEQ2511	ACTAATCAAAC-AACTCATCAA----AGTTACTAT-AATAG--TAGCACTCCTGCTAGT

**Table 25: Comparative Sequences relating to SAG0368
(protein of unknown function)**

SEQ2501	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2502	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2503	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2504	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2505	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2506	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2507	ACTATAGCAGTAACAC-TAACACAGGTCAGGCTGATTCAAGTGGAAGTGTTAATAATTA
SEQ2508	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2509	ACTATAGCAGTAACAC-TAACACAGGTCAGGCTGATTCAAGTGGAAGTGTTAATAATTA
SEQ2510	GTTGAACCAGGGACACATAAAATAAATGGAGAACAAGCACTTGTTTATTCTCGTATGCG
SEQ2511	ACTATAGCAGTAACAC-TAACACAGGTCAGGCTGATTCAAGTGGAAGTGTTAATAATTA
SEQ2501	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2502	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2503	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2504	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2505	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2506	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2507	AACGGGGCTGCAACGCCTAATCCAAACACAGGAACGCAACCAGTACCAGGTCAAACCTAA
SEQ2508	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2509	AACGGGGCTGCAACGCCTAATCCA-----
SEQ2510	TATGATGATCCAGAGGGAGATTATGGGCGTCAAAAAAGACAACGTGAAGTAATTCAAAA
SEQ2511	AACGGGGCTGCAACGCCTAATCCAAACACAGGAACGCAACCAGTACCAGGTCAAACCTAA
SEQ2501	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2502	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2503	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2504	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2505	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2506	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2507	CCA-----
SEQ2508	GTCCTTAAAAAATATTGGCGTTAAATAGTA-----
SEQ2509	-----
SEQ2510	GTCCTTAAAAAATATTGGCGTTAAATAGTATTAGTTCATACAAAAAATTCCTTCCGC
SEQ2511	CCA-----
SEQ2501	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2502	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2503	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2504	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2505	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2506	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	GTAAGTAATAACATGCAAACCTAATATTGAGATATCATCAAAAACGATTCCTAATTTGTT
SEQ2511	-----
SEQ2501	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2502	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAAGGTGAAGACGCTAC
SEQ2503	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2504	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2505	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2506	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	GCTTATAAAGATTCATTGGAACATATTAAATCTTATCAGTTGAAGGGTGAAGACGCTAC
SEQ2511	-----

**Table 25: Comparative Sequences relating to SAG0368
(protein f unknown function)**

SEQ2501	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2502	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2503	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2504	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2505	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2506	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	TTATCAGATGGTGGCTCTTATCAAATTTTAACTAAGAAACATCTACTTGCAGTTCAAAA
SEQ2511	-----
SEQ2501	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2502	AGAATTAAGAAAGAACTAGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2503	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2504	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2505	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2506	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	AGAATTAAGAAAGAACTGGATAAAAAGCGTAGTAAACTCTGAAGACAAGCGCGATTCT
SEQ2511	-----
SEQ2501	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2502	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2503	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2504	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2505	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2506	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	TATGAAGATTACTATGGTACTACTGCTAGTAATGATTCTTCTACTTATTCATCAACACA
SEQ2511	-----
SEQ2501	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2502	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2503	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2504	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2505	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2506	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	GAGAATAATTATAATACAACACCTTATTCAGAAGCACCACCAAGTTACAGTGGTAATAC
SEQ2511	-----
SEQ2501	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2502	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAAGTTACTATAATAGTAGCACTC
SEQ2503	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2504	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2505	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2506	ACTTATTAGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2507	-----
SEQ2508	-----
SEQ2509	-----
SEQ2510	ACTTAT-AGTTCTGAGACTAATCAAACAACCTCATCAAATTACTATAATAGTAGCACTC
SEQ2511	-----

**Table 25: Comparative Sequences relating to SAG0365
(protein of unknown function)**

SEQ2501 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2502 TGCTAATAACTATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2503 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2504 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2505 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2506 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2507 -----
 SEQ2508 -----
 SEQ2509 -----
 SEQ2510 TGCTAGTAACATATAGCAGTAACACTAACACAGGTCAGGCTGATTCAAGTGGAAGTGTCA
 SEQ2511 -----

SEQ2501 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2502 TAATCATAATGGGGCTGCAACGCCTAATCCA
 SEQ2503 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2504 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2505 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2506 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2507 -----
 SEQ2508 -----
 SEQ2509 -----
 SEQ2510 TAATCATAACGGGGCTGCAACGCCTAATCCA
 SEQ2511 -----

>SEQ ID NO 2550: 54_090 frame: 1
 YNFSTNELSKTFKDFKLAQSKSHAIEETKPFISILLMGVDTGSEHRKSKWSGNSDSMILVT
 INPKTNKTTMTSLERDVLKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDILLDINV
 DYFMQINMQGLVDLVNAVGGITVTNKFDFFISIAANEPEYKAVVEPGTHKINGEQALVYS
 RMRYDDPEGDYGRQKRQREVIQVLKILALNSISSYKKILSAVSNMOTNIEISSKTIP
 NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
 AILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
 STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2551: 54_1169NT frame: 1
 YNFSTNELSKTFKDFKLAQSKSHAIEETKPFISILLMGVDTGSEHRKSKLVRK.RFYDLSH
 YKS.N..NNDKLR.T.RID.IEWSQK.WTDWRRSKAKCSLCFWWCNGNIDDCSRLIRY.C
 .LLYAN.YARIS.FSQCCWYNSN..I.LSNINCCQ.TRVQGCC.TRDT.NKWRSTCLF
 SYAL..SRGLWASKKTT.SNSKSP.KNIGVK.Y.FIQKNSFRSK..HAN.Y.DIIKND
 .FVSL.RFIGTY.ILSVER.RRYFIRWWLLSNEN.ETSTCSSK.N.ERTR.KA..NSEDK
 RDSI.RLLWYYC...FFYLFINTRE.L.YNTLFRSTTKLOW.YYL.F.D.SNNSSKLL..
 .HSC..L.Q.H.HRSG.FKWKQ.C.S.WGCNA.S

>SEQ ID NO 2552: 54_18RS21 frame: 1
 YNFSTNELSKTFKDFKLAQSKSHAIEETKPFISILLMGVDTGSEHRKSKWSGNSDSMILVT
 INPKTNKTTMTSLERDVLKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDILLDINV
 DYFMQINMQGLVDLVNAVGGITVTNKFDFFISIAANEPEYKAVVEPGTHKINGEQALVYS
 RMRYDDPEGDYGRQKRQREVIQVLKILALNSISSYKKILSAVSNMOTNIEISSKTIP
 NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
 AILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
 STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2553: 54_2603 frame: 1
 YNFSTNELSKTFKDFKLAQSKSHAIEETKPFISILLMGVDTGSEHRKSKWSGNSDSMILVT
 INPKTNKTTMTSLERDVLKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDILLDINV
 DYFMQINMQGLVDLVNAVGGITVTNKFDFFISIAANEPEYKAVVEPGTHKINGEQALVYS
 RMRYDDPEGDYGRQKRQREVIQVLKILALNSISSYKKILSAVSNMOTNIEISSKTIP
 NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
 AILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
 STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2554: 54_A909 frame: 1
 YNFSTNELSKTFKDFKLAQSKSHAIEETKPFISILLMGVDTGSEHRKSKWSGNSDSMILVT
 INPKTNKTTMTSLERDVLKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDILLDINV
 DYFMQINMQGLVDLVNAVGGITVTNKFDFFISIAANEPEYKAVVEPGTHKINGEQALVYS
 RMRYDDPEGDYGRQKRQREVIQVLKILALNSISSYKKILSAVSNMOTNIEISSKTIP
 NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS

**Table 25: Comparative Sequences relating to SAG0360
(protein of unknown function)**

AIIYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2555:54_CJB110 frame: 1
YNFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
INPKTNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
DYFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
RMRYDDPEGDYGRQKRQREVIQVKKILALNSISSYKKILSAVSNNMQTNIEISSKTIP
NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
AIIYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTY.F.D.SNNSKLL..

>SEQ ID NO 2556:54_COH1 frame: 1
DFKLDKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTNKTTMTSL
ERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINVDFYFMQINMQGLVD
LVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGR
QKRQREVIQVKKILALNSISSYKKILSAVSNNMQTNIEISSKTIPNLLAYKDSLEHIK
SYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTSAILYEDYYGTTAS
NDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTY.F.D.SNNSKLL...HSC..L.Q.H.H
RSG.FKWKC..L.RGCNA.SKHRNATSTRSN.S

>SEQ ID NO 2557:54_H36B frame: 1
YNFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
INPKTNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
DYFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
RMRYDDPEGDYGRQKRQREVIQVKKILALNSISSYKKILSAVSNNMQTNIEISSKTIP
NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
AIIYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2558:54_JM9130013 frame: 1
YNFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
INPKTNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
DYFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
RMRYDDPEGDYGRQKRQREVIQVKKILALNSISSYKKILSAVSNNMQTNIEISSKTIP
NLLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTCKHLLAVQNRIKKELDKKRSKTLKTS
AIIYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
STPASNYSSNTNTGQADSSGSVNNHNGAATPNP

>SEQ ID NO 2559:54_M781 frame: 2
SILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTNKTTMTSLERDVLIKLSGPKNNGQTG
VEAKLNAAYASGGAEMALMTVQDLLDINVDFYFMQINMQGLVDLVNAVGGITVTNKFDFP
SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGRQKRQREVIQVKKILAL
NSISSYKKILSAVSNNMQTNIEISSKTIPNLLAYKDSLEHIKSYQLKGEDATLSDGGSYQ
ILTCKHLLAVQNRIKKELDKKRSKTLKTSAILYEDYYGTTASNDSSSTYSSTQENNYNTTP
YSEAPPSYSGNTTYSSETNQTHQSYNSSTPASNYSSNTNTGQADSSGSVNNYNGAATP
NPNTGTQPVPGQTNP

SEQ2550	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2551	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKLVRKRYDLSHY
SEQ2552	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2553	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2554	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2555	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2556	-----DFKLDKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2557	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2558	NFSTNELSKTFKDFKLAKSKSHAIEETKPFSSILLMGVDTGSEHRKSKWSGNSDSMILVT
SEQ2559	-----SILLMGVDTGSEHRKSKWSGNSDSMILVT

SEQ2550	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2551	SNNDKLRTRIDIEWSQKWTDWRRS-----KAKCSLCFWWCGNGIDDCSRLIRYCLLY
SEQ2552	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2553	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2554	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2555	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2556	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV
SEQ2557	NPKNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAYASGGAEMALMTVQDLLDINV

**Table 25: Comparative Sequences relating to SAG0368
(protein of unknown function)**

SEQ2558	NPKTNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAAYASGGAEMALMTVQDLLDINV
SEQ2559	NPKTNKTTMTSLERDVLIKLSGPKNNGQTGVEAKLNAAAYASGGAEMALMTVQDLLDINV
SEQ2550	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2551	NYARISFSQCCWWYNS-----NILSNINCCQTRVQGCCTRDNTNKTSTCLFSY
SEQ2552	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2553	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2554	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2555	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2556	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2557	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2558	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2559	YFMQINMQGLVDLVNAVGGITVTNKFDFPISIAANEPEYKAVVEPGTHKINGEQALVYS
SEQ2550	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2551	LSRGRWLWASKKTTSNSKSPKNIGVKYFIQKNSFRSKHANYDI IKNDSFVSLRFIGTYI-
SEQ2552	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2553	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2554	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2555	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2556	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2557	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2558	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2559	MRYDDPEGDYGRQKRQREVIQKVLKKILALNSISSYKKILSAVSNMOTNIEISSKTIP
SEQ2550	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2551	L-SVERRRYFIRWLLSNFNETSTCSSKNERTRKANSEDKRDSIRLLWYCYFFYLFINT
SEQ2552	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2553	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2554	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2555	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2556	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2557	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2558	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2559	LLAYKDSLEHIKSYQLKGEDATLSDGGSYQILTKKHLLAVQNRIKKELDKKRSKTLKTS
SEQ2550	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2551	ELY-----NTLFRST-----TKLQWYYLFDSNNSSKLLHSCLOH
SEQ2552	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2553	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2554	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2555	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2556	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2557	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2558	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQNYNS
SEQ2559	ILYEDYYGTTASNDSSSTYSSTQENNYNTTPYSEAPPSYSGNTTYSSETNQTHQSYNS
SEQ2550	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2551	RSGFKWKQSWGCNAS-----
SEQ2552	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2553	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2554	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2555	-----
SEQ2556	RSGFKWKCLRGCNASKHRNATSTRSNS-----
SEQ2557	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2558	TPASNYSSNTNTGQADSSGSVNNHNGAATPNP-----
SEQ2559	TPASNYSSNTNTGQADSSGSVNNHNGAATPNPNTGTQPVPGQTNP

Table 26: Comparative Sequences relating to SAG0503 (lipase/acylhydrolase)

**SEQ ID NO. 2601: SAG0503 FROM THE 090 GBS TYPE Ia STRAIN
(REVERSE COMPLEMENT)**

GGGCACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAA
AGACTTCCTAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATAC
AACCTCTCAAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAA
TTATGGTGTGTCTGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGA
GAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTTCGTAAAGAGCTCAGTCATTTATC
ACTAAATTCCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAGAAATACTTGCAAAGCAAGACAAGATAA
TCCTAAATTGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAATAAATGCAAAAC
CGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGA
CCGCTTTATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGC
TCTCTTACTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
TGAAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGTAACAAG

**SEQ ID NO. 2602: SAG0503 FROM THE H36b GBS TYPE Ib STRAIN
(REVERSE COMPLEMENT)**

TTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCCT
AACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCA
AGGTGGTTTTGTTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTGT
GTCGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTGA
TTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTTCGTAAAGAGCTCAGTCATTTATCACTAAATTC
TTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAGAAATCCTTGCAAAGCAAGACAAGATAATCCTAAATT
GCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAATAAATGCAAACCGTTATTGA
TAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTTA
TAAGGGAATAAATGGTAAAGAGGGTATTATAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTAC
TGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAAG
AAAAAAGTGGCCGAACCCAGCTTTCTTGTAACAAGTGCTC

**SEQ ID NO. 2603: SAG0503 FROM THE 18RS21 GBS TYPE II STRAIN (REVERSE
COMPLEMENT)**

GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCC
TAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTC
AAGGTGGTTTTGTTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTG
TGTCTGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTG
ATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTTCGTAAAGAGCTCAGTCATTTATCACTAAATT
CCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAGAAATCCTTGCAAAGCAAGACAAGATAATCCTAAAT
TGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAATAAATGCAAACCGTTATTG
ATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTT
ATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTA
CTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAA
GAAAAAAGTGGCCGAACCCAGCTTTCTTGTAACA

**SEQ ID NO. 2604: SAG0503 FROM THE COH1 GBS TYPE III STRAIN (REVERSE
COMPLEMENT)**

GGACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAG
ACTTCCTAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGGGATACAA
CCTCTCAAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATT
ATGGTGTGTCTGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGA
AAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTTCGTAAAGAGCTCAGTCATTTATCAC
TAAATTCCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAGAAATCCTTGCAAAGCAAGACAAGATAATC
CTAAATTGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAATAAATGCAAACCG
TTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACC
GCCTTTATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTC
TCTTACTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATG
AAACAAGAAAAAAGTGGCCGAACCCAGCTTTCTTGTAACAA

Table 26: Comparative Sequences relating to SAG0503 (lipase/acylhydrolase)

SEQ ID NO. 2605: SAG0503 FROM THE CJB110 GBS NONTYPEABLE STRAIN (REVERSE COMPLEMENT)

GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCC
TAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTC
AAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTG
TGCTCGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTG
ATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATT
CCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAAGAAATACTTGCAAAAGCAAGACAAGATAATCCTAAAT
TGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAATGCAAACCGTTATTG
ATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTT
ATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTA
CTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAA
GAAAAAAGCTGGCCGAACCCAGCTTTCTTGACAA

SEQ ID NO. 2606: SAG0503 FROM THE 1169NT1 GBS TYPE V STRAIN (REVERSE COMPLEMENT)

GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCC
TAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTC
AAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTG
TGCTCGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTG
ATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATT
CCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAAT
TGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAATGCAAACCGTTATTG
ATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTT
ATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTA
CTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAA
GAAAAAAGCTGGCCGAACCCAGCTTTCTTGACAA

SEQ ID NO. 2607: SAG0503 FROM THE JM9130013 GBS TYPE VIII STRAIN (REVERSE COMPLEMENT)

GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCC
TAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTC
AAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTG
TGCTCGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTG
ATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATT
CCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAAT
TGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAATGCAAACCGTTATTG
ATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTT
ATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTA
CTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAA
GAAAAAAGCTGGCCGAACCCAGCTTTCTTGACAA

SEQ ID NO. 2608: SAG0503 FROM THE 2603 V/R GBS TYPE V STRAIN (REVERSE COMPLEMENT)

AGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAGACTTCC
CTAACAAAGAAAGTTATCCCACTTAACATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTC
CAAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGGTG
GTGTCTGGGAATACTAGTCAACAAATTTTAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTTAGAGAAAGCTG
GATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGCTGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAAT
TCCTTTGAGAAACCAGCAGAAGCATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAAT
TTGCCTATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAATGCAAACCGTTATTG
GATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAATGTTTATTTTGTCCCAATTAATGACCGCCTTT
TATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTA
ACTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAATGAAACAA
AGAAAAAAGCTGGCCGAACCCAGCTTTCTTGACAAAGTGG

Table 26: Comparative Sequences relating to SAG0503 (lipase/acylhydrolase)

SEQ ID NO. 2609: SAG0503 FROM THE M781 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)

GGACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAAATCCTAAATTAACAAAAAAG
ACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGCTCTTGGAGATTCTCTGACCGAAGGTGTGGGGGATACAA
CCTCTCAAGGTGGTTTTGTCCCACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTAAAT
ATGGTGTGCTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGAAAAAGATTAGAGA
AAGCTGATTTATTGACGCTAACTGTTGGTGAATGATGTCTTGGCTGTTATTCGTAAAGAGCTCAGTCATTATCAC
TAAATTCCTTTGAGAAACCAGCAGAAAGCATATAAGGAACGTTTGAAAGAAATTCCTGCAAAAGCAAGACAAGATAATC
CTAAATTCCTTATTTATGTTTTAGGCATTTATAATCCTTTTACCTAACTTCCACAATTAATAAAATGCAAAACCG
TTATTGATAATTGGAATAAAGCTACAAAAGAGTAGTTGATGCTTCAGAAAATGTTATTTTGTCCCAATTAATGACC
GCCTTTATAAGGGAATAAATGGTAAAGAGGGTATTACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTC
TCTTTACTGGAGACCATTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAATG
AAACAAGAAAAAAGCTGGCCGAACCCAGCTTTCTTGATACAA

SEQ2601 GGCACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2602 -----TTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2603 -----GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2604 GGACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2605 -----GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2606 -----GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2607 -----GTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2608 -----AGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA
SEQ2609 GGACAAGTTTGTACAAAAAGCAGGCTCTATTTTTTCCTTGATCATTCCAAAATCAA

SEQ2601 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2602 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2603 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2604 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2605 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2606 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2607 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2608 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC
SEQ2609 TCCTAAATTAACAAAAAAGACTTCCTAACAAAGAAAGTTATCCCACTTAATATGTTGC

SEQ2601 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2602 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2603 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2604 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2605 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2606 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2607 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2608 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC
SEQ2609 TCTTGGAGATTCTCTGACCGAAGGTGTGGGCGATACAACCTCTCAAGGTGGTTTTGTCCC

SEQ2601 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2602 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2603 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2604 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2605 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2606 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2607 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2608 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG
SEQ2609 ACTGCTATCAGAATCACTCCATAATCGATACTCTTACCAAGTGACTTCTGTTAATTATGG

SEQ2601 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2602 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2603 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2604 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2605 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2606 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2607 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2608 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA
SEQ2609 TGTGTCTGGGAATACTAGTCAACAAATTTTAAAACGTATGACGACAGATCCTCAAATCGA

SEQ2601 AAAAGATTTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGC
SEQ2602 AAAAGATTTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGC
SEQ2603 AAAAGATTTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCTTGGC

Table 26: Comparative Sequences relating to SAG0503 (lipase/acylhydrolase)

SEQ2604	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2605	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2606	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2607	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2608	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2609	AAAAGATTAGAGAAAGCTGATTTATTGACGCTAACTGTTGGTGGTAATGATGTCCTTGGC
SEQ2601	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2602	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2603	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2604	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2605	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2606	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2607	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2608	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2609	TGTTATTCGTAAAGAGCTCAGTCATTTATCACTAAATTCCTTTGAGAAACCAGCAGAAGC
SEQ2601	ATATAAGGAACGTTTGAAAGAAATACTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2602	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2603	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2604	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2605	ATATAAGGAACGTTTGAAAGAAATACTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2606	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2607	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2608	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2609	ATATAAGGAACGTTTGAAAGAAATCCTTGCAAAAGCAAGACAAGATAATCCTAAATTGCC
SEQ2601	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2602	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2603	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2604	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2605	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2606	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2607	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2608	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2609	TATTTATGTTTTAGGCATTTATAATCCTTTTTACCTAAACTTTCCACAATTAACATAAAAT
SEQ2601	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2602	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2603	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2604	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2605	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2606	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2607	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2608	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2609	GCAAACCGTTATTGATAATTGGAATAAAGCTACAAAAGAAGTAGTTGATGCTTCAGAAAA
SEQ2601	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2602	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2603	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2604	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2605	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2606	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2607	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2608	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2609	TGTTTATTTTGTCCCAATTAATGACCGCCTTTATAAGGGAATAAATGGTAAAGAGGGTAT
SEQ2601	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2602	TATAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2603	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2604	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2605	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2606	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2607	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2608	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2609	TACAGAGTCATCAAATAGTCAGGCAAGTATCACTAATGATGCTCTCTTTACTGGAGACCA
SEQ2601	TTTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAAATAAA
SEQ2602	TTTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAAATAAA
SEQ2603	TTTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAAATAAA

Table 26: Comparative Sequences relating to SAG0503 (lipase/acylhydrolase)

SEQ2604 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
 SEQ2605 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
 SEQ2606 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
 SEQ2607 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
 SEQ2608 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA
 SEQ2609 TTTTCATCCCAATAATATTGGCTATCAAATCATGTCTAACGCCGTTATGGAGAAAATAAA

SEQ2601 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAAAG-----
 SEQ2602 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAAAGTGGTCC-----
 SEQ2603 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAA-----
 SEQ2604 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAA-----
 SEQ2605 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAA-----
 SEQ2606 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAA-----
 SEQ2607 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAA-----
 SEQ2608 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAAAGTGG-----
 SEQ2609 TGAACAAGAAAAAACTGGCCGAACCCAGCTTTCTTGACAAATABCMARATVSTNCSRA

SEQ2601 -----
 SEQ2602 -----
 SEQ2603 -----
 SEQ2604 -----
 SEQ2605 -----
 SEQ2606 -----
 SEQ2607 -----
 SEQ2608 -----
 SEQ2609 NGTSAGASACYHYDAS

>SEQ ID NO 2650:103_090 frame: 2

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVP
 LSESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLA
 VIRKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKM
 QTVIDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDH
 FHPNNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2651:103_H36B frame: 2

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLS
 ESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIR
 KELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKMQTV
 IDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDHFHP
 NNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2652:103_18RS21 frame: 3

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLS
 ESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIR
 KELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKMQTV
 IDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDHFHP
 NNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2653:103_COH1 frame: 3

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVPL
 LSESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAV
 IRKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKMQ
 TVIDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDHF
 HPNNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2654:103_CJB110 frame: 3

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLS
 ESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIR
 KELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKMQTV
 IDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDHFHP
 NNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2655:103_1169NT frame: 3

IFSLIIPKSNPKLTKKDFTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLS
 ESLHNRYSYQVTSVNYGVSGNTSQQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIR
 KELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPIYVLGIYNPFYLNFPQLTKMQTV
 IDNWNKATKEVVDASENVYFVPINDRLYKINGKEGITESSNSQASITNDALFTGDHFHP
 NNIGYQIMSNVMEKINETRKNWP

Table 26: Comparativ Sequences relating to SAG0503 (lipase/acylhydrolase)

>SEQ ID NO 2656:103 JM9130013 frame: 3

IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLS
ESLHNRYSYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIR
KELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTV
IDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFH
NNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2657:103 2603 frame: 1

IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLL
SESLHNRYSYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVI
RKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQT
VIDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFH
PNNIGYQIMSNVMEKINETRKNWP

>SEQ ID NO 2658:103 M781 frame: 3

IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPL
LSESLHNRYSYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAV
IRKELSHLSLNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQ
TVIDNWNKATKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFH
HPNNIGYQIMSNVMEKINETRKNWP

SEQ2650	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2651	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2652	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2653	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2654	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2655	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2656	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2657	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY
SEQ2658	IFSLIIPKSNPKLTKKDFLTKKVIPLNYVALGDSLTEGVGDTSQGGFVPLLSSESLHNRY

SEQ2650	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2651	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2652	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2653	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2654	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2655	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2656	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2657	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS
SEQ2658	SYQVTSVNYGVSGNTSQILKRMTTDPQIEKDLEKADLLTLTVGGNDVLAVIRKELSHLS

SEQ2650	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2651	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2652	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2653	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2654	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2655	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2656	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2657	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA
SEQ2658	LNSFEKPAEAYKERLKEILAKARQDNPKLPYVVLGIYNPFYLNFPQLTKMQTVIDNWNKA

SEQ2650	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2651	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2652	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2653	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2654	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2655	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2656	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2657	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI
SEQ2658	TKEVVDASENVYFVPINDRLYKGINGKEGITESSNSQASITNDALFTGDHFFHPNNIGYQI

SEQ2650	MSNAVMEKINETRKNWP
SEQ2651	MSNAVMEKINETRKNWP
SEQ2652	MSNAVMEKINETRKNWP
SEQ2653	MSNAVMEKINETRKNWP
SEQ2654	MSNAVMEKINETRKNWP
SEQ2655	MSNAVMEKINETRKNWP
SEQ2656	MSNAVMEKINETRKNWP
SEQ2657	MSNAVMEKINETRKNWP
SEQ2658	MSNAVMEKINETRKNWP

**Table 27: C mparative Sequences relating to SAG1473
(cell wall surface anchor family pr tein)**

**SEQ ID NO. 2701: SAG1473 FROM THE 1169NT1 GBS TYPE V STRAIN
(REVERSE COMPLEMENT)**

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CGTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGA
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
GTGATGGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2702: SAG1473 FROM THE 18RS21 GBS TYPE II STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CGTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGA
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2703: SAG1473 FROM THE 2603 V/R GBS TYPE V STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CGTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGA
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2704: SAG1473 FROM THE 090 GBS TYPE Ia STRAIN

GACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATCCGTC
AACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGA
AGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTATTAAG
AATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGA
TGGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2705: SAG1473 FROM THE A909 GBS TYPE Ia STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATTAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2706: SAG1473 FROM THE CJB110 GBS NONTYPEABLE STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CGTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGA
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

**Table 27: C mparativ Sequences relating to SAG1473
(cell wall surface anchor family protein)**

**SEQ ID NO. 2707: SAG1473 FROM THE COH1 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)**

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGGAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGAACGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2708: SAG1473 FROM THE H36b GBS TYPE Ib STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATTAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAArAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2709: SAG1473 FROM THE JM910013 GBS TYPE VIII STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATTAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGTAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2710: SAG1473 FROM THE M732 GBS TYPE III STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGGAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGAACGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ ID NO. 2711: SAG1473 FROM THE M781 GBS TYPE III STRAIN

GATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAAAGATCAGATGA
ACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCATCAAGTGAACCAGAAACAAATC
CCTCAACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGATGGGAGC
ACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATTAATTTTCAGAAGATAGTAT
TAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAA
ATGATGAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA

SEQ2701	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2702	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2703	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2704	-----
SEQ2705	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2706	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2707	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2709	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2710	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA
SEQ2711	ATACAAGTGATAAGAATACTGACACGAGTGTCTGACTACGACCTTATCTGAGGAGAAA

**Table 27: C mparative Sequences relating to SAG1473
(cell wall surface anch r family pr tein)**

SEQ2701	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2702	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2703	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2704	-----GACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2705	GATTAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2706	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2707	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCA
SEQ2709	GATTAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCGAGTTCA
SEQ2710	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCA
SEQ2711	GATCAGATGAACTAGACCAGTCTAGTACTGGTTCTTCTTCTGAAAATGAATCAAGTTCA
SEQ2701	TCAAGTGAACCAGAAACAAATCCGTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2702	TCAAGTGAACCAGAAACAAATCCGTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2703	TCAAGTGAACCAGAAACAAATCCGTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2704	TCAAGTGAACCAGAAACAAATCCGTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2705	TCAAGTGAACCAGAAACAAATCCCTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2706	TCAAGTGAACCAGAAACAAATCCGTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2707	TCAAGTGAACCAGAAACAAATCCCTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2709	TCAAGTGAACCAGAAACAAATCCCTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2710	TCAAGTGAACCAGAAACAAATCCCTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2711	TCAAGTGAACCAGAAACAAATCCCTCAACTAATCCACCTACAACAGAACCATCGCAACCC
SEQ2701	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2702	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2703	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2704	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2705	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2706	TCACCTAGTGAAGAGAACAAGCCTGATGGTAGAACGAAGACAGAAATTGGCAATAATAAG
SEQ2707	TCACCTAGTGAAGAGAACAAGCCTGATGGGAGCACGAAGACAGAAATTGGCAATAATAAG
SEQ2709	TCACCTAGTGAAGAGAACAAGCCTGATGGGAGCACGAAGACAGAAATTGGCAATAATAAG
SEQ2710	TCACCTAGTGAAGAGAACAAGCCTGATGGGAGCACGAAGACAGAAATTGGCAATAATAAG
SEQ2711	TCACCTAGTGAAGAGAACAAGCCTGATGGGAGCACGAAGACAGAAATTGGCAATAATAAG
SEQ2701	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2702	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2703	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2704	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2705	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2706	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2707	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2709	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2710	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2711	GATATTTCTAGTGGAAACAAAAGTATTAATTTTCAGAAGATAGTATTAAGAATTTTAGTAAA
SEQ2701	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2702	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2703	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2704	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2705	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2706	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2707	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2709	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2710	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT
SEQ2711	GCAAGTAGTGATCAAGAAGAAGTGGATCGCGATGAATCATCATCTTCAAAGCAAATGAT

SEQ2701 GGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----

**Table 27: Comparative Sequences relating t SAG1473
(cell wall surface anchor family protein)**

SEQ2702	GGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2703	GGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2704	GGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2705	GAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2706	GGGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2707	GAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAASGATACAAGTGATAAGAATACTGACAC
SEQ2709	GAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2710	GAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAA-----
SEQ2711	GAGAAAAAAGGCCACAGTAAGCCTAAAAAGGAATABCMARATVSTNCSRATNGTSAGCWA
SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	AGTGTCGTGACTACGACCTTATCTGAGGAGAAAAGATTAGATGAACTAGACCAGTCTAG
SEQ2709	-----
SEQ2710	-----
SEQ2711	TRACANCHRAMYRTN-----
SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	ACTGGTTCTTCTTCTGAAAATGAATCGAGTTCATCAAGTGAACCAGAAACAAATCCCTC
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----
SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	ACTAATCCACCTACAACAGAACCATCGCAACCCTCACCTAGTGAAGAGAACAAGCCTGA
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----
SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	GGTAGCACGAAGACAGAAATTGGCAATAATAAGGATATTTCTAGTGGAACAAAAGTATT
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----
SEQ2701	-----

**Table 27: Comparative Sequences relating to SAG1473
(cell wall surface anchor family protein)**

SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	ATTTCAGAAGATAGTATTAAGAATTTTAGTAAAGCAAGTAGTGATCAAGAARAAGTGGA
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----

SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	CGCGATGAATCATCATCTTCAAAAGCAAATGATGAGAAAAAAGGCCACAGTAAGCCTAA
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----

SEQ2701	-----
SEQ2702	-----
SEQ2703	-----
SEQ2704	-----
SEQ2705	-----
SEQ2706	-----
SEQ2707	AAGGAA
SEQ2709	-----
SEQ2710	-----
SEQ2711	-----

>SEQ ID NO 2750:4_1169NT frame: 1
 DTS DKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
 SPSEENKPDGR TKTEIGNNKDISSG TKVLISED SIKNFSKASSDQEEVDRDESSSSKASD
 GKKGH SKPKKE

>SEQ ID NO 2751:4_18RS21 frame: 1
 DTS DKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
 SPSEENKPDGR TKTEIGNNKDISSG TKVLISED SIKNFSKASSDQEEVDRDESSSSKAND
 GKKGH SKPKKE

>SEQ ID NO 2752:4_2603 frame: 1
 DTS DKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
 SPSEENKPDGR TKTEIGNNKDISSG TKVLISED SIKNFSKASSDQEEVDRDESSSSKAND
 GKKGH SKPKKE

>SEQ ID NO 2753:4_090 frame: 1
 DQSSTGSSSENESSSSSEPETNPSTNPPTTEPSQPSPSEENKPDGR TKTEIGNNKDISSG
 TKVLISED SIKNFSKASSDQEEVDRDESSSSKANDGKKGH SKPKKE

>SEQ ID NO 2754:4_A909 frame: 1
 DTS DKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
 SPSEENKPDGR TKTEIGNNKDISSG TKVLISED SIKNFSKASSDQEEVDRDESSSSKAND
 EKKGH SKPKKE

>SEQ ID NO 2755:4_CJB110 frame: 1
 DTS DKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
 SPSEENKPDGR TKTEIGNNKDISSG TKVLISED SIKNFSKASSDQEEVDRDESSSSKAND
 GKKGH SKPKKE

>SEQ ID NO 2756:4_COH1 frame: 1

**Table 27: Comparative Sequences relating to SAG1473
(cell wall surface anchoring family protein)**

DTSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVERDESSSSKAND
EKKGHSKPKKE

>SEQ ID NO 2757:4_H36B frame: 1

DTSDKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEXVDRDESSSSKAND
EKKGHSKPKKE

>SEQ ID NO 2758:4_JM9130013 frame: 1

DTSDKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
EKKGHSKPKKE

>SEQ ID NO 2759:4_M732 frame: 1

DTSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVERDESSSSKAND
EKKGHSKPKKE

>SEQ ID NO 2760:4_M781 frame: 1

DTSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
EKKGHSKPKKE

SEQ2750	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2751	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2752	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2753	-----DQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2754	TSDKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2755	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2756	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2757	TSDKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2758	TSDKNTDTSVVTTLSEEKRLDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2759	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP
SEQ2760	TSDKNTDTSVVTTLSEEKRSDELQSSSTGSSSENESSSSSEPETNPSTNPPTTEPSQP

SEQ2750	SPSEENKPDGRKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKASD
SEQ2751	SPSEENKPDGRKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2752	SPSEENKPDGRKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2753	SPSEENKPDGRKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2754	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2755	SPSEENKPDGRKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2756	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVERDESSSSKAND
SEQ2757	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEXVDRDESSSSKAND
SEQ2758	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND
SEQ2759	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVERDESSSSKAND
SEQ2760	SPSEENKPDGSTKTEIGNNKDISSGTVLISEDSIKNFSKASSDQEEVDRDESSSSKAND

SEQ2750	KKGHSKPKKE
SEQ2751	KKGHSKPKKE
SEQ2752	KKGHSKPKKE
SEQ2753	KKGHSKPKKE
SEQ2754	KKGHSKPKKE
SEQ2755	KKGHSKPKKE
SEQ2756	KKGHSKPKKE
SEQ2757	KKGHSKPKKE
SEQ2758	KKGHSKPKKE
SEQ2759	KKGHSKPKKE
SEQ2760	KKGHSKPKKE

**Tabl 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

**SEQ ID NO. 2801: SAG1552 FROM THE 1169NT1 GBS TYPE V STRAIN
(REVERSE COMPLEMENT)**

TTTGTGTTAAAGGTGATACTGTACTTCACAAGCCCACCAATAAACCTTTTGTGTTAAAGGAGTAGACGTTGAGTCT
TCCTTAGCAGGTTATCATCACAAACGATTTTCCCTATTACTCAAAAAACGTATCGTGAGTGGTTCCATTTAATTTCCAAC
ATGGGGGCAAATACTGTAAGAGTCAAAGTACCGATGAATGTTGCATTTTACGATGCTTTATATCACCACAACAAAGCA
TCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCTATAACAGCTTTTAAT
GATAATTATAGGGGGTATTTAAACGAGAAGCAAAAGGCGTTGTGGATATTCTCCATGGGCGTAAGCAAGTATGGAAT
ACTGATTTTGGTAGCCGTCATTATCATTATGATCTTAGTCCCTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAAT
AGTGGTACTGTCGCTTATACTAATCATCAAGAGAAAAAACGCAATATAAAGGACGTTATTTTAAACCTCTCGCGCA
GCTAATCCATTTGAGGTCATGCTAGCTCAAGTTATGGATGAATTGACACATTATGAGACAGCTAAATATGGTTGGCAA
CATTGTATTAGTTTTTCAAACCTCACCAACACAGACCCTTTTCGTTATCGAAAACCATTGAGGCACAGGCTCCTAAA
TACGTACAACATAAATGTAGAAAATATTCAAGCTAATTCGAATGTTAAAGCAGGTATTTTGCAGCATATAAAGCTATT
GATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAGAATATCAGTAAAGAAGATAGACAAAAGATTAAA
GAACTTTCTTTGTACAGGGATACGTTAAACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGC
TATTCGACAGCGAGAGGTATTGCCCAAAAGAAATTGATAAACGTCCTCTGCCGATTAAATGAAAAGAACAAAGGTACAG
CGTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTTGGAGCGACTATCAATGCATGGCAAGACGATTGG
AATGCAAGGGCGTGAATACATCCTTCGCCACAAATAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAAT
CAAGGTTATGGTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCAAGGAGAGTGG
AAACATCCTCTG

SEQ ID NO. 2802: SAG1552 FROM THE

ATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAGTGATGAAAGCTATCTCTACCTTGCGATTAAACAAAACCT
GAAAACTAAAAGAAAAACGATTATTACCAATAGATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAG
GTCACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATTATTTGTCCAAGAGCGC
TATAATGCCTTAAAAGCGAATCTCTTCGACAGCTTAACGGTAAAGATTTTATGCTTTCCACCAAGAAAGAACAGT
AGTAATTTTGAGCAGATCAATATGGTATTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAGCAACAGAG
AGGTTCTTACCAACTCATCCTACTGGTCTTCTCAAAACAGGAACAATTGATAGGCACCAAAAAACATTTGATTCACAA
ACAGATATTTCTGTTTGGAAAGGACTTTATAGAGGTCAGAAATCCGTGGCAGTTGTTGAATTTTCTGATCCCATCATCT
CAAAAAATTACGATGATTACTTTAAACATTATGGTGTGAAGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGT
GCTAATAGCAAAGAAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACC
TTTTTAAAGACTCCTATTATAGTATTTAAGAAAGA

SEQ ID NO. 2803: SAG1552 FROM THE 18RS21 GBS TYPE II STRAIN

AAGGGCTTATTAAAGAAAAATACAAGAACTAACTTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCACCAATAAA
CCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCTTAGCGGGTTATCATCACACGATTTTCCCTATTACTCAAAAA
ACGTATCGTGAATGGTTCCATTTAATTTCCAACATGGGGGCAATACTGTAAGAGTCAAGGTACCGATGAATGTTGCA
TTTACGATGCCTTATATCACCACAACAAAGCATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCT
TATCGCAATAATGCTTCTATAACAGCTTTTAATGATAATTATAGGGGGTATTTAAACGAGAAGCAAAAGGCGTTGTG
GATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGGTAGCCGTCATTATCATTATGATCTTAGTCCCTGG
GTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGTACTGTGCTTATACTAATCATCAAGAGAAAAAACGCAA
TATAAAGGACGTTATTTTAAACCTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGATGAATTG
ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTCAAACTCACCAACAACAGACCCTTTTCAT
TATCGAAAACCATTTGAGGCACAGGCTCCTAAATACGTACCAATAATGTAGAAAATATTCAAGCTAATTCAAATGTT
AAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAG
AATATCAGTAAAGAAGATAGACAAAAGATTAAAGAACTTCTTTGTACAGGGATACGTTAAACTGCTAAATGCTTAT
CACAAAATCCCTGTTCTAGTCACGGGTTATGGCTATTTCGACAGCGAGAGGTATTGCCCAAAAGAAATTGATAAACGT
CCTCTGCCGATTAATGAAAAGAAACAAGGTACGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGAATACATCTTTCGCCACAAATAAACATAGT
CAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTATGGTTTATTAGGCTTTAAAAACGCAAAACATCATTAT
CAAGTTGATGGTAAAGAGGCAAGGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCT
AGCAGTGATGAAAGCTATCTCTACCTTGCGATTAAACAAAACCTGAAAACTAAAAGAAAAACGATTATTACCAATA
GATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTATTG
TCTATTGATCCAAATGGCAAGTCTGAATTATTTGTCCAAGAGCGCTATAATGCCTTAAAGCGAACTATCTTCGACAG
CTTAACGGTAAAGATTTTTTATGCTTTCCACCAAGAAAGAACAGTAGTAATTTTGAAGCAGATAAATATGGTATTGAGA
AATACAAAGATTGTTGAAGACATGGAAAAAGTAAAGCAACAGAGAGGTTCTTACCAACTCAATATCTGGTCTTCTC
AAAAACAGGAACAACCTGATAGGCACCAAAAAACATTTGATTACAAACAGATATTTTCGTTTGGAAAGGACTTTATAGAG
GTCAGAATTCGGTGGCAGTTGTTGAATTTTCTGATCATCATCTCAAAAAATTACAGATGATTACTTTAAACATTAT
GGTGTGAAGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAACACACTGATAAAGATG
GCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACCTTTTTAAAGACTCCTATTATGTATTAAGAAAG
AA

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ ID NO. 2804: SAG1552 FROM THE 2603 V/R GBS TYPE V STRAIN
(REVERSE COMPLEMENT)

TATTAAAAGAAAATACAAGAACTAACTTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCACCAATAAACCTTTTG
TTGTTAAAGGAGTAGACGTTGAGTCTTCCTTAGCGGGTTATCATCACAACGATTTTCCTATTACTCAAAAAACGTATC
GTGAATGGTTCCATTTAATTTCCAACATGGGGCAAATACTGTAAGAGTCAAGGTACCGATGAATGTTGCATTTTACG
ATGCCTTATATCACCACAACAAAGCATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCA
ATAATGCTTCTATAACAGCTTTTAATGATAATTATAGGGGGTATTTAAACGAGAAGCAAAGCGTTGTGGATATTCT
TCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGGTAGCCGTCATTATCATTTATGATCTTAGTCCTTGGGTACTTG
GTTATGTCGTAGGGGATGATTGGAAATAGTGGTACTGTCGCTTATACTAATCATCAAGAGAAAAAACGCAATATAAAG
GACGTTATTTTAAACTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGATGAATTGACACATT
ATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACTCACCAACAACAGACCCCTTTTCATTATCGAA
AACCATTTGAGGCACAGGCTCCTAAATACGTACAACCTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCAG
GTATGTTTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAGAATATCA
GTAAAGAAGATAGACAAAAGATTAAAGAACCTTTCTTGTACAGGGATACGTTAAACTGCTAAATGCTTATCACA
TCCCTGTTCTAGTCACGGGTTATGGCTATTCGACAGCGAGGTTATGCCCAAAAAGAAATTGATAAACGTCCTCTGC
CGATTAATGAAAAAGAACAAAGGTGAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTTTGGAGCGA
CTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCTTTCGCCACAAATAAACATAGTCAATTCC
TATGGGGGGATGCACAAGTATTTAATCAAGGTTATGGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTG
ATGGTAAAAGAGGCAAAGGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAGTG
ATGAAAGCTATCTCTACCTTGGGATTAAAAACAAACCTGAAAACTAAAAGAAAAACGATTATTACCAATAGATATTA
CACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTG
ATCCAAATGGCAAGTCTGAATTATTTGTCCAAGAGCGCTATAATGCCTTAAAGCGAACTATCTTCGACAGCTTAACG
GTAAAGATTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGTATTGAGAAATACAA
AGATTGTTGAAGACATGAAAAAGTAAAGCAACAGAGAGGTTCTTACCAACTCATCCTACTGGTCTTCTCAAAACAG
GAACAACTGATAGGCACCAAAAAACATTTGATTACAAACAGATATTTTCGTTTGGAAAGGACTTTATAGAGGTGAGAA
TTCCGTGGCAGTTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACGATGATTACTTTAAACATTATGGTGTGA
AGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAAACACACTGATAAAGATGGCAGATT
ATCGTTTGA AAAATTGGGAGAGACCCGATACCAAAACCTTTTAAAGACTCCTATTATAGTATTAAGAAAGAAATGGT
CTAAAGAAAGAGAGAGAACATATGGTCCA

SEQ ID NO. 2805: SAG1552 FROM THE A909 GBS TYPE Ia STRAIN
(REVERSE COMPLEMENT)

AAGGGCTTATTAAAAGAAAATACAAGAACTAACTTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCACCAATAAA
CCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCCTTAGCGGGTTATCATCACAACGATTTTCCTATTACTCAAAAA
ACGTATCGTGAATGGTTCCATTTAATTTCCAACATGGGGCAAATACTGTAAGAGTCAAGGTACCGATGAATGTTGCA
TTTTACGATGCCTTATATCACCACAACAAAGCATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCT
TATCGCAATAATGCTTCTATAACAGCTTTTAATGATAATTATAGGGGGTATTTAAACGAGAAGCAAAGGCGTTGTG
GATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGGTAGCCGTCATTATCATTATGATCTTAGTCCTTGG
GTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGTACTGTCGCTTATACTAATCATCAAGAGAAAAAACGCAA
TATAAAGGACGTTATTTTAAACTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGATGAATTG
ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCAACAACAGACCCCTTTTCAT
TATCGAAAACCAATTTGAGGCACAGGCTCCTAAATACGTACAACCTAAATGTAGAAAATATTCAAGCTAATTCAAATGTT
AAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAG
AATATCAGTAAAGAAGATAGACAAAAGATTAAAGAACCTTTCTTGTACAGGGGATACGTTAAACTGCTAAATGCTTAT
CACAAAATCCCTGTTCTAGTCACGGGTTATGGCTATTCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGT
CCTCTGCCGATTAATGAAAAAGAACAAAGGTGAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTTT
GGAGCGACTATCAATGGCATGGCAAGACGATTGGAATGCAAGGGCGTGAATACATCTTTCGCCACAAATAAACATAGT
CAATTCTTATGGGGATGCACAAGTATTTAATCAAGGTTATGGTTTATTAGGCTTTAAAAACGCAAAACATCATTAT
CAAGTTGATGGTAAAAGAGGCAAAGGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCT
AGCAGTGATGAAAGCTATCTCTACCTTGGGATTAAAAACAAACCTGAAAACTAAAAGAAAAACGATTATTACCAATA
GATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTATTG
TCTATTGATCCAAATGGCAAGTCTGAATTATTTGTCCAAGAGCGCTATAATGCCTTAAAGCGAACTATCTTCGACAG
CTTAACGGTAAAGATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGTATTGAGA
AATACAAAGATTGTTGAAGACATGAAAAAGTAAAGCAACAGAGAGGTTCTTACCAACTCATCCTACTGGTCTTCTC
AAAACAGGAACAACATGATAGGCACCAAAAAACATTTGATTACAAACAGATATTTTCGTTTGGAAAGGACTTTATAGAG
GTCAGAAATCCGTGGCAGTTGTTGAATTTTCTGATCCATCATCTCAAGAATTACGATGATTACTTTAAACATTAT
GGTGTGAAGGAGTTAGAAAATTGAGAGCCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAAACACACTGATAAAGA
TGGCAGATTATCGTTTGA AAAATTGGGAGAGACCCGATACCAAAACCTTTTAAAGA

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ ID NO. 2806: SAG1552 FROM THE CJB110 GBS NONTYPEABLE STRAIN

TATTACTTTGATGGTAGTTTGTATTTACCAAAGGGCTTATTTAAAGAAAATACAAGAACTAACTTTGTTGTTAAAGGT
GATACTGTACTTCACAAGCCCACCAATAAACCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCCTTAGCGGGTTAT
CATCACAACGATTTTCTATTACTCAAAAACGATATCGTGAATGGTTCCATTTAATTTCCAACATGGGGGCAAATACT
GTAAGAGTCAAGGTACCGATGAATGTTGCATTTTACGATGCCTTATATCACCACAACAAAGCATCAAAGAGGCCACTG
TATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCTATAACAGCTTTTAATGATAATTATAGGGGG
TATTTAAAACGAGAAGCAAAGGCGTTGTGGATATTCTCCATGGGCGTAAGCAAGTATGGAATACAGATTTTGGTAGC
CGTCATTATCATTATGATCTTAGTCCTTGGGTACTTGGTTATGTGCTAGGGGATGATTGGAATAGTGGTACTGTCGCT
TATACTAATCATCAAGAGAAAAAACGCAATATAAAGGACGTTATTTTAAACTTCTGTGGCAGCTAATCCATTTGAG
GTCATGCTAGCTCAAGTAATGGATGAATTGACACATTTATGAGCAGCTAAATATGGTTGGCAACATTTGATTAGTTTT
TCAAACCTACCAACACAGACCCCTTTTCATTATCGAAAACCATTTGAGGCACAGGCTCCTAAATACGTACAACATAAT
GTAGAAAATATTTCAAGCTAATTCAAATGTTAAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCCATCCTCGA
TACAAGGATTATCTATTATTGATAAAGAGAATATCAGTAAAGAAGATAGACAAAAGATTAAAGAACTTTCTTTGTCA
CAGGGATACGTTAAACTGCTAATGCTTATCACAATAATCCCTGTTCTAGTCACGGGTATGGCTATTCCGACAGCGAGA
GGTATTGCCCAAAAAGAAATGATAAACGTCCTCTGCCGATTAATGAAAAAGAAACAAGGTCAGCGTTTACTAGAAGAT
TATGAATCTTTTATATCATCCGGTAGTTTGGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGG
AATACATCTTTCCGCCACAAATAACATAATCAATTCCTATGGGGGAGTGCACAAGTATTTAATCAAGGTTATGGTTTA
TTAGGCTTTAAAAACGCAAAACATCATTAAGTTGATGGTAAAAGAGGCAAAGGAGAGTGGAAACATCCTCTGATG
ACTAGTGCACAGGAGATGACTTATATGCTAGCAGTGATGAAAGCTATCTCTACCTTGCATTAAAAACAAAACCTGAA
AACTAAAAGAAAAACGATTATTACCAATAGATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTC
ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATTTATTGTCCAAGAGCGCTAT
AATGCCTTTAAAGCGAACTATCTTCGACAGCTTAACGGTAAAGATTTTATGCTTTCCACCAAAGAAGACAGTAGT
AATTTTGAGCAGATAAATATGGTATTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGG
TTCTTACCAACTCATCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATTTGATTCACAAACA
GATATTTTCGTTTGGAAAGGACTTTATAGAGGTGAGAATCCGTGGCAGTTGTTGAATTTTCTGATCCATCATCTCAA
AAAATTACGATGATTACTTTAAACATTATGGTGTGAAGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGTGCT
AATAGCAAAGAAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACCTTT
TTAAAGACTCCTATTATGTATTAAGAAAGA

SEQ ID NO. 2807: SAG1552 FROM THE COH1 GBS TYPE III STRAIN

TTTACCACAGGGCTTATTTAAAGAAAATACAAGAACTAACTTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCAC
CAATAAACCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCCTTAGCGGGTTATCATCACAACGATTTTCTTATTAC
TCAAAAACGATATCGTGAATGGTTCCATTTAATTTCCAACATGGGGGCAAATACTGTAAGAGTCAAGGTACCGATGAA
TGTTGCATTTTACGATGCCCTATATCACCACAACAAAGAATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTAT
AGATTCCTATCGCAATAATGCTTCTATAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAACGAGAAGCAAAGG
CGTTGTGGATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTTGGTAGCCGTCATTATCATTATGATCTTAG
TCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGTACTGTGCTTATACTAATCATCAAGAGAAAAA
AACGCAATATAAAGGACGTTATTTTAAACTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGA
TGAATTGACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCAACAACAGACCC
TTTTCATTTATCGAAAACCATTTGAGGCACAGGCTCCTAAATACGTACAACATAATGTAGAAAATATTCAAGCTAATTC
AAATGTTAAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATATCTATTATTGTA
TAAAGAGAATATCAGTAAAGAAGATAGACAAAAGATTAAAGAACCTTCTTTGTCACAGGGATACGTTAAACTGCTAAA
TAAACGTCCTCTGCCGATTAATGAAAAGAACAGGTGAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGG
TAGTTTTGGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGAATACATCTTTGCCACAAATAA
ACATAGTCAATTCTATGGGGGATGCACAAGTATTTAATCAAGGTTATGGTTTATTAGGCTTTAAAAACGCAAAACA
TCATTATCAAGTTGATGGTAAAAGAGGCAAAGGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTT
ATATGCTAGCAGTGATGAAAGCTATCTCTACCTTGGCATTAAACAAAACCTGAAAACTAAAAGAAAACGATTATT
ACCAATAGATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTACATTTTCTAAATCTAGTGACTT
TGTATTGTCTATTGATCCAAATGGCAAGTCTGAATTTTGTCCAAGAGCGCTATAATGCCTTAAAGCGAACTATCT
TCGACAGCTTAACGGTAAAGATTTTATGCTTTCCACCAAAGAAGAAGAGTAGTAATTTTGAAGCAGATAAATATGGT
ATTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTTACCAACTCATCTACTGG
TCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATTTGATTACCAACCAGATATTTCTGTTTGGAAAGGACTT
TATAGAGGTGAGAATCCGTGGCAGTTGTTGAATTTTCTGATCCATCATCTCAAAAATTACAGATGATTACTTTAA
ACATTATGGTGTGAAGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAACACACTGAT
AAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACCTTTTTTAAAGACT

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ ID NO. 2808: SAG1552 FROM THE H36b GBS TYPE Ib STRAIN

AAGGGGCTTATTAAGAAATACAAGAACTAATTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCAACATAA
ACCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCCTTAGCGGGTATCATCACAAACGATTTTCTATTACTCAAAA
AACGTATCGTGAATGGTTCCATTTAATTTCCAACATGGGGGCAAATACTGTAAGAGTCAAGGTACCGATGAATGTTGC
ATTTTACGATGCCTTATATCACCACAACAAAGCATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATT
TTATCGCAATAATGCTTCTATAACAGCTTTAATGATAATTATAGGGGGTATTTAAACGAGAAGCAAAAGGCGTTGT
GGATATTCTCCATGGGCGTAAGCAAGTATGGAATCTGATTTTGGTAGCAGTCATTATCATTATGATCTTAGTCCCTG
GGTACTTGTTATGTCGTAGGGGATGATGGACATAGTGGTACTGTGCGCTTTATACTAATCATCAAGAGGAGAAAAACG
CAATATAAAGGACGTTATTTTAAACTTCTGTGGCAGCTAATCCATTGTAGGTCATGCTAGCTCAAGTAATGGATGAA
TTGACACATTATGAGACAGCTAAATATGGTTGGCAACATTGTATTAGTTTTCAAACTCACCAACAACAGACCCTTTT
CATTATCGAAAACCATTTGAGGCACAGGCTCCTAAATACGTACAACATAATGTAGAAAATATTCAAGCTAATTCGAAT
GTTAAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAA
GAGAATATCAGTAAAGAGATAGACAAAAGATTAAAGAACTTTCTTTGTACAGGGATACGTTAAACTGCTAAATGCT
TATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTACTCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAA
CGTCTCTGCGGATTAATGAAAAAGAACAAAGGTGAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGT
TTTGGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGTGTGGAATACATCCTTCGCCACAAATAACAT
AGTCAATTCCTATGGGGGGATGCACAAGTATTTAATCAAGGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATAT
TATCAGGTTGATGGTAAAGAGGCAAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATAT
GCTAGCAGTGATGAAAGCTATCTACCTTGCGATTAAAAACAAACCTGAAAACTAAAAGAAAAACGATTATTACCA
ATAGATATTACACCAAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTA
TTGTCTATTGATCCAAATGGCAAGTCTGAATTATTTGTCCAAGAGCGCTATAACGCCTTAAAAGCGAATATCTTCGA
CAGCTTAATGGTAAAGATTTTTATGCTTTCCACCAAGAAAGACAGTAGTAATTTGAGCAGATAAATATGGTATTG
AGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAGCAACAGAGAGGTCTTACCAACTCATCCTACTGGTCTT
CTCAAAACAGGAACAACTGATAGGCACCAAAAACATTTGATTCACAAACAGATATTTCTGTTTGGAAAGGACTTTATA
GAGGTGAGAATCCGTGGCAGTTGTTGAATTTTCTGATCCATCATCTCAAAAAATTCACGATGATTACTTTAACAT
TATGGTGTGAAGGAGTTAGAAATTGAGAGCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAACACACTGATAAAG
ATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACCTTTTAAAAGACTCCTATTATAGT

SEQ ID NO. 2809: SAG1552 FROM THE JM9130013 GBS TYPE VIII STRAIN

ACTTTGTTGTTAAAGGTGATACTGTACTTCACAAGCCCAACATAAACCTTTTGTGTTAAAGGAGTAGACGTTGAGT
CTTCTTAGCGGGTATCATCACAAACGATTTTCTATTACTCAAAAACGTATCGTGAATGGTTCCATTTAATTTCCA
ACATGGGGGCAAATACTGTAAGAGTCAAGGTACCGATGAATGTTGCATTTTACGATGCCTTATATCACCACAACAAAG
CATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCTATAACAGCTTTTA
ATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAAGGCGTTGTGGATATTCTCCATGGGCGTAAGCAAGTATGGA
ATACTGATTTTGTAGCAGTCATTATCATTATGATCTTAGTCTTGGGACTTGGTTATGTCGTAGGGGATGATTGGA
ATAGTGGTACTGTGCGCTTATACTAATCATCAAGAGAAAAACGCAATATAAAGGACGTTATTTTAAACTTCTGTGG
CAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGATGAATTGACACATTATGAGACAGCTAAATATGGTTGGC
AACATTTGATTAGTTTTTCAAACCTCACCAACAACAGACCCCTTTTATTATCGAAAACCATTTGAGGCACAGGCTCCTA
AATACGTACAATAATGTAGAAAAATTTCAAGCTAATTCGAATGTTAAAGCAGGTATGTTTGCAGCATATAAAGCTA
TTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAGAATATCAGTAAAGAAGATAGACAAAAGATTA
AAGAACCTTTCTTTGTACAGGGATACGTTAAACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATG
GCTACTCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGAAAAAGAACAAAGGTC
AGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTTGGAGCGACTATCAATGCATGGCAAGACGATT
GGAATGCAAGGGTGTGGAATACATCCTTCGCCACAAATAAACATAGTCAATTCCATGAGGGGATGCACAAGTATTTA
ATCAAGGTTATGGTTTATTAGGCTTTAAAACGCAAAACATCATTATCAGGTTGATGGTAAAGAGGCAAGAGAGT
GGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAGTGATGAAAGCTATCTCTACCTTGCGA
TTAAAACAAAACCTGAAAACTAAAAGAAAAACGATTATTACCAATAGATATTACACCAAAATCTGGTAGTAGAAAAA
TGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTATTGCTATTGATCCAAATGGCAAGTCTGAATTAT
TTGTCCAAGAGCGCTATAACGCCTTAAAAGCGAATATCTTCGACAGCTTAATGGTAAAGATTTTTATGCTTTCCAC
CAAAGAAGAACAGTAGTAATTTGAGCAGATAAATATGGTATGAGAAATACAAAGATTGTTGAAGACATGGAAAAAG
TAAAAGCAACAGAGAGGTTCTTACCAACTCATCCTACTGGTCTTCTCAAAACAGGAACAACCTGATAGGCACCAAAAA
CATTTGATTACAAACAGATATTTCTGTTTGGAAAGGACTTTATAGAGGTCAGAATCCGTGGCAGTTGTTGAATTTTT
CTGATCCATCATCTCAAAAAATTCACGATGATTACTTTAACATTATGGTGTGAAGGAGTTAGAAATTGAGAGCATTG
CTTTAGGATTAGGTGCTAATAGCAAAGAAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGAC
CCGATACCAAAACCTTTTAAAAGACTCCTATTATAGTATTAAGAAAG

**Table 28: Comparativ Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ ID NO. 2810: SAG1552 FROM THE M732 GBS TYPE III STRAIN

TACAAGAACTAAGTTTGTGTTAAAGGTGATACTGTACTTCACAAGCCACCAATAAACCTTTTGTGTTAAAGGAGT
AGACGTTGAGTCTTCTTAGCGGGTTATCATCACAACGATTTTCTATTACTCAAAAAACGTATCGTGAATGGTTCCA
TTTAATTTCCAACATGGGGGCAATACTGTAAGAGTCAAGGTACCGATGAATGTTGCATTTTACGATGCCTTATATCA
CCACAACAAAGAAATCAAAGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCTAT
AACAGCTTTTAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTGGATATTCTCCATGGGCGTAA
GCAAGTATGGAATACTGATTTTGGTAGCCGTCATTATCATTATGATCTTAGTCTTGGGTACTTGGTTATGTCGTAGG
GGATGATTGCAATAGTGGTACTGTGCTTATACTAATCATCAAGAGAAAAAACGCAATATAAAGGACGTTATTTTAA
AACTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTAATGGATGAATTGACACATTATGAGACAGCTAA
ATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCAACACAGACCCCTTTTATTATCGAAAACCATTTGAGGC
ACAGGCTCCTAAATACGTAACTAATGTAGAAATATTCAGCTAATTCAAATGTTAAAGCAGGTATGTTTGCAGC
ATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATAAAGAGAATATCAGTAAAGAAGATAG
ACAAAAGATTAAAGAACTTTCTTTGTCACAGGGGATACGTTAACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGT
CACGGGTTATGGCTATTTCGACAGCGAGAGGTATTGCCCAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGAAAA
AGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTTGGAGCGACTATCAATGCGATG
GCAAGACGATTGGAATGCAAGGGCGTGGAAATACATCTTTCGCCACAAATAAACATAGTCAATTCCTATGGGGGATGC
ACAAGTATTTAATCAAGGTTATGGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAGAGG
CAAAGGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAGTGATGAAAGCTATCT
CTACCTTGCGATTAAACAAAACCTGAAAACTAAAAGAAAAACGATTATTACCAATAGATATTACACCAAAATCTGG
TAGTAGAAAAATGAATGGTAGTAAGGTCACATTTTCTAAATCTAGTGACTTTGTATTGTCATTGATCCAAATGGCAA
GTCTGAATTATTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAATCTCTTCGACAGCTTAACGGTAAAGATTTTAA
TGCTTTCCACCAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGTATTGAGAAATACAAAGATTGTTGAAGA
CATGGAAAAAGTAAAAGCAACAGAGAGGTTCTTACCAACTCATCCTACTGGTCTTCTCAAAACAGGAACAACCTGATAG
GCACCAAAAAACATTTGATTACAAACAGATATTTCTGTTTGGAAAGGACTTTATAGAGGTCAGAATTCGGTGGCAGTT
GTTGAATTTTTCTGATCCATCATCTCAAAAAATTCACGATGATTACTTTAAACATTATGGTGTGAAGGAGTTAGAAAT
TGAGAGCATTGCTTTAGGATTAGGTGCTAATAGCAAAGAAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAA
TTGGGAGAGACCCGATACCAAAACCTTTTTTAAAAGACTCCTATTATAGTATTAAG

SEQ ID NO. 2811: SAG1552 FROM THE M781 GBS TYPE III STRAIN

TTTGATGGTAGTTTGTATTTACCACAGGGCTTATTAAGAAAAATACAAGAACTAAGTTTGTGTTAAAGGTGATACT
GTACTTCACAAGCCCAACATAAACCTTTTGTGTTAAAGGAGTAGACGTTGAGTCTTCTTAGCGGGTTATCATCAC
AACGATTTTCTTATTACTCAAAAAACGTATCGTGAATGGTTCCATTTAATTTCCAACATGGGGGCAATACTGTAAGA
GTCAAGGTACCGATGAATGTTGCATTTACGATGCCTTATATCACCACAACAAAGAATCAAAGAGGCCACTGTATTTG
TTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCTATAACAGCTTTTAAATGATAATTATAGGGGGTATTTA
AAACGAGAAGCAAAAGGCGTTGTGGATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTTGGTAGCCGTCAT
TATCATTATGATCTTAGTCTTGGGTACTTGGTTATGCTTACGGGATGATTGGAATAGTGGTACTGTGCTTATACT
AATCATCAAGAGAAAAAACGCAATATAAAGGACGTTATTTTAAAACCTCTGTGGCAGCTAATCCATTTGAGGTGATG
CTAGCTCAAGTAAATGGATGAATTGACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAAC
TCACCAACAACAGACCCCTTTTATTATCGAAAACCATTTGAGGCACAGGCTCCTAAATACGTACAACATAATGTAGAA
AATATTCAAGCTAATTCAAATGTTAAAGCAGGTATGTTTGCAGCATATAAAGCTATTGATTTCATCCTCGATACAAG
GATTATCTATTATTTGATAAAGAGAATATCAGTAAAGAAGATAGACAAAAGATTAAAGAACTTCTTTGTACAGGGGA
TACGTTAAACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTACAGGTTATGGCTATTTCGACAGCGAGAGGTATT
GCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAAATGAAAAAGAACAGGTGAGCGTTTACTAGAAGATTATGAA
TCTTTTATATCATCCGGTAGTTTTTGGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAAATACA
TCTTTTCGCCACAAATAAACATAGTCAATTCCTATGGGGGATGCACAAGTATTTAATCAAGGTTATGGTTTATTAGGC
TTTAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCAAGGAGAGTGGAAACATCCTCTGATGACTAGT
GCAACAGGAGATGACTTATATGCTAGCAGTGATGAAAGCTATCTCTACCTTGCATTAAAACAAAACCTGAAAACTA
AAAGAAAAACGATTATTACCAATAGATATTACACCAAAATCTGGTAGTAAAAATGAATGGTAGTAAGGTCACATTT
TCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATTATTTGTCCAAAGAGCGCTATAATGCC
TAAAAGCGAATATCTTCGACAGCTTAACGGTAAAGATTTTATGCTTTCCACCAAGAAGAAGAGTAGTAATTTT
GAGCAGATAAATATGGTATTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTTA
CCAATCATCCTACTGGTCTTCTCAAAACAGGAACAACGATAGGCACCAAAAAACATTTGATTACAAACAGATATT
TCGTTTGGAAAGGACTTTATAGAGGTCAGAATTCGGTGGCAGTTGTTGAATTTTTCTGATCCATCATCTCAAAAAAT
CACGATGATTACTTTAAACATTATGGTGTGAAGGAGTTAGAAATGAGAGCATTGCTTTAGGATTAGGTGCTAATAGC
AAAGAAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACCAAAACCTTTTTAAAA
GACTCCTATTATAGTATTAAGAAAGAATGG

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

[illegible]

**Tabl 28: Comparative Sequences relating to SAG1552
(conserved hypothetical pr tein)**

SEQ2809	AGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCT
SEQ2810	AGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCT
SEQ2811	AGAGGCCACTGTATTTGTTGCAAGGAATACGTATAGATTCTTATCGCAATAATGCTTCT
SEQ2801	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2802	-----
SEQ2803	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2804	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2805	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2806	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2807	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2808	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2809	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2810	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2811	TAACAGCTTTTAAATGATAATTATAGGGGGTATTTAAAACGAGAAGCAAAGGCGTTGTG
SEQ2801	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2802	-----ATGACTA-GTGCAACAGGAGATGACTTATAT-GCTAGCAGTGATGAAAGC
SEQ2803	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2804	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2805	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2806	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2807	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2808	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2809	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2810	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2811	ATATTCTCCATGGGCGTAAGCAAGTATGGAATACTGATTTGGTAGCCGTCAATTATCAT
SEQ2801	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2802	TAT--CTCTA--CCTTGCG-ATTAAACAAAACCTGAAAACTAAAAAGAAAACGATTAT
SEQ2803	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2804	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2805	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2806	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2807	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2808	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2809	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2810	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2811	TATGATCTTAGTCCTTGGGTACTTGGTTATGTCGTAGGGGATGATTGGAATAGTGGT-AC
SEQ2801	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2802	TACCAATAGATATTA--CACCAAAATCTGGTAGTAAAAATGAATGGTAGTAAGGTCAC
SEQ2803	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2804	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2805	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2806	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2807	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2808	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2809	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2810	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2811	TGTCGCTT-ATACTAATCATCAAGAGAA-AAAAACGCAATATAAAGGAC-GTTATTTTAA
SEQ2801	AACCTTCTGCGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2802	ATTTTCTAAATCTAGTGA-CTTTGTATTGTC-TATTGATCCAAATGGCAAGTCTGAATT-
SEQ2803	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2804	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2805	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2806	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2807	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2808	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2809	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2810	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2811	AACCTTCTGTGGCAGCTAATCCATTTGAGGTCATGCTAGCTCAAGTATGGAT--GAATTG
SEQ2801	ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2802	ATTTGTC-CAAGAGCGCTATA-ATGCCT--TAAAAGCGAACTATCTTCGACAGCTTAACG
SEQ2803	ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2804	ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA

Table 28: Comparative Sequences relating to SAG1552 (conserved hypothetical protein)

SEQ2805 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2806 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2807 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2808 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2809 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2810 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA
SEQ2811 ACACATTATGAGACAGCTAAATATGGTTGGCAACATTTGATTAGTTTTTCAAACCTACCA

SEQ2801 CAACAGAC-----CCTTTTCGTTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2802 TAAAGATTTTTTATGCTTTCCACCAAAGAAGACAGTAGTAATTTTGAGCAGATCAATA
SEQ2803 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2804 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2805 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2806 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2807 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2808 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2809 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2810 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA
SEQ2811 CAACAGAC-----CCTTTTCATTATCGAAAACCATTTG-AGGCACAGGCTCCTAAATA

SEQ2801 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTGCAATGTTAAAGCA---GGTATTT
SEQ2802 GGTATTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGT
SEQ2803 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2804 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2805 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2806 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2807 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2808 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2809 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2810 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT
SEQ2811 G-TACAACTAAATGTAGAAAATATTCAAGCTAATTCAAATGTTAAAGCA---GGTATGT

SEQ2801 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2802 TCTTACCAACTCATCCTACTGGTCTT---CTCAAAACAGGAACAATTGAT-AGGCACCA
SEQ2803 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2804 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2805 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2806 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2807 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2808 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2809 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2810 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA
SEQ2811 TTGCAGCATATAAAGCTATTGATTTCCATCCTCGATACAAGGATTATCTATTATTTGATA

SEQ2801 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2802 AAAAACTTTTGATTACAAAACAGATATTTCTGTTTGGAAAGGACTTTATAGAGGTGAGAAT
SEQ2803 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2804 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2805 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2806 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2807 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2808 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2809 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2810 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA
SEQ2811 AAGAGAATATCAGTAAAGAAGATAGACAAAAGATT-AAAGAACTTTCTTTGTACAGGGA

SEQ2801 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2802 TCCGTGGCAGTTGTTGAATTTTTCTGATCCA---TCATCTCAAAAATTCACGATGATTA
SEQ2803 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2804 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2805 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2806 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2807 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2808 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2809 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2810 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA
SEQ2811 TACGTTA-AACTGCTAAATGCTTATCACAAAATCCCTGTTCTAGTCACGGGTTATGGCTA

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ2801	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2802	TTTAAACATTATGGTGTGAAGGAGTTAGAAATTGA-GAGCATTGCTTTAGGATTAGGTG
SEQ2803	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2804	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2805	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2806	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2807	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2808	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2809	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2810	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2811	TCGACAGCGAGAGGTATTGCCCAAAAAGAAATTGATAAACGTCCTCTGCCGATTAATGA
SEQ2801	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2802	TAATAGCAAAGAAAACACACTGATAAAGATGGCAGAT-----TATCGTTTGAAAAATT
SEQ2803	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2804	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2805	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2806	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2807	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2808	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2809	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2810	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2811	AAAGAACAAGGTCAGCGTTTACTAGAAGATTATGAATCTTTTATATCATCCGGTAGTTT
SEQ2801	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2802	GGAGAGAC--CCGATAC----CAAACCTTTTAA-----AAGACTCCTATTATAGTATT
SEQ2803	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2804	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2805	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2806	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2807	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2808	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2809	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2810	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2811	GGAGCGACTATCAATGCATGGCAAGACGATTGGAATGCAAGGGCGTGGAATACATCCTT
SEQ2801	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2802	A--AGAAAGAA-----
SEQ2803	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2804	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2805	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2806	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2807	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2808	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2809	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2810	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2811	GCCACAAATAAACATAGTCAATTCTATGGGGGGATGCACAAGTATTTAATCAAGGTTA
SEQ2801	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2802	-----
SEQ2803	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2804	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2805	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2806	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2807	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2808	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2809	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2810	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2811	GGTTTATTAGGCTTTAAAAACGCAAAACATCATTATCAAGTTGATGGTAAAAGAGGCCAA
SEQ2801	GGAGAGTGGAACATCCTCTG-----
SEQ2802	-----
SEQ2803	GGAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2804	GGAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2805	GGAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2806	GGAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2807	GGAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2808	GAAGAGTGGAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG

**Table 28: Comparative Sequences relating to SAG1552
(conserv d hypothetical protein)**

SEQ2809	GAAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2810	GGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2811	GGAGAGTGGAAACATCCTCTGATGACTAGTGCAACAGGAGATGACTTATATGCTAGCAG
SEQ2801	-----
SEQ2802	-----
SEQ2803	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2804	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2805	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2806	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2807	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2808	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2809	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2810	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2811	GATGAAAGCTATCTCTACCTTGGCATTAAAACAAAACCTGAAAACTAAAAGAAAAACG
SEQ2801	-----
SEQ2802	-----
SEQ2803	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2804	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2805	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2806	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2807	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2808	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2809	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2810	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2811	TTATTACCAATAGATATTACACCAAATCTGGTAGTAGAAAAATGAATGGTAGTAAGGT
SEQ2801	-----
SEQ2802	-----
SEQ2803	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2804	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2805	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2806	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2807	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2808	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2809	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2810	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2811	ACATTTTCTAAATCTAGTGACTTTGTATTGTCTATTGATCCAAATGGCAAGTCTGAATT
SEQ2801	-----
SEQ2802	-----
SEQ2803	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2804	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2805	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2806	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2807	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2808	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2809	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2810	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2811	TTTGTCCAAGAGCGCTATAATGCCTTAAAAGCGAACTATCTTCGACAGCTTAACGGTAA
SEQ2801	-----
SEQ2802	-----
SEQ2803	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2804	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2805	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2806	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2807	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2808	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2809	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2810	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2811	GATTTTTATGCTTTCCACCAAAGAAGAACAGTAGTAATTTTGAGCAGATAAATATGGT
SEQ2801	-----
SEQ2802	-----
SEQ2803	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2804	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

SEQ2805	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2806	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2807	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2808	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2809	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2810	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2811	TTGAGAAATACAAAGATTGTTGAAGACATGGAAAAAGTAAAAGCAACAGAGAGGTTCTT
SEQ2801	-----
SEQ2802	-----
SEQ2803	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2804	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2805	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2806	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2807	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2808	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2809	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2810	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2811	CCAACTCATCCTACTGGTCTTCTCAAACAGGAACAACCTGATAGGCACCAAAAAACATT
SEQ2801	-----
SEQ2802	-----
SEQ2803	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2804	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2805	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2806	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2807	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2808	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2809	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2810	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2811	GATTACAAACAGATATTTTCGTTTGAAAGGACTTTATAGAGGTCAGAATTC CGTGGCA
SEQ2801	-----
SEQ2802	-----
SEQ2803	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2804	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2805	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2806	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2807	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2808	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2809	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2810	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2811	TTGTTGAATTTTCTGATCCATCATCTCAAAAAATTACCGATGATTACTTTAAACATTA
SEQ2801	-----
SEQ2802	-----
SEQ2803	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2804	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2805	GGTGTGAAGGAGTTAGAAATTGAGAGCCATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2806	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2807	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2808	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2809	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2810	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2811	GGTGTGAAGGAGTTAGAAATTGAGAG--CATTGCTTTAGGATTAGGTGCTAATAGCAAA
SEQ2801	-----
SEQ2802	-----
SEQ2803	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2804	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2805	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2806	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2807	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2808	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2809	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2810	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC
SEQ2811	AAAACACACTGATAAAGATGGCAGATTATCGTTTGAAAAATTGGGAGAGACCCGATACC

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical pr tein)**

```

SEQ2801 -----
SEQ2802 -----
SEQ2803 AAACCTTTTAAAAGACTCCTATTATGTATTAAGAAAGAA-----
SEQ2804 AAACCTTTTAAAAGACTCCTATTATAGTATTAAGAAAGAATGGTCTAAAGAAAGAGAG
SEQ2805 AAACCTTTTAAAAGA-----
SEQ2806 AAACCTTTTAAAAGACTCCTATTATGTATTAAGAAAGA-----
SEQ2807 AAACCTTTTAAAAGACT-----
SEQ2808 AAACCTTTTAAAAGACTCCTATTATAGT-----
SEQ2809 AAACCTTTTAAAAGACTCCTATTATAGTATTAAGAAAG-----
SEQ2810 AAACCTTTTAAAAGACTCCTATTATAGTATTAAG-----
SEQ2811 AAACCTTTTAAAAGACTCCTATTATAGTATTAAGAAAGAATGG-----

```

```

SEQ2801 -----
SEQ2802 -----
SEQ2803 -----
SEQ2804 GAACATATGGTCCA
SEQ2805 -----
SEQ2806 -----
SEQ2807 -----
SEQ2808 -----
SEQ2809 -----
SEQ2810 -----
SEQ2811 -----

```

>SEQ ID NO 2850:62_1169NT frame: 1

FVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFLHLSNMGANTVRV
KVPNMVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLKREAGVVD
ILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKKQYKGRYFKTS
AAANPFEVMLAQVMDDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPFQAQPKYVQLNV
ENIQANSNVKAGIFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLSQGYVKLLNA
YHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFISSGSFGATINAW
QDDWNARAWNTSFATNKSQFLWGDAQVFNQGYGLLGFKNAKHYYQVDGKRGKGEWKHPL
MTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMGSKVTFKSSD
FVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINMVLNRNTKIV
EDMEKVKATERFLPTHPTGLLKTGTIDRHQKTFDSQTDISFGKDFIEVRIPWQLNFSDF
SSQKIHDYFKHYGVKELEIESIALGLGANSKENTLIKMDYRLKNWERPDTKTFKDSY
YSI.ER

>SEQ ID NO 2851:62_18RS21 frame: 1

KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFLH
ISNMGANTVRVVKPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRG
YLKREAGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKK
QYKGRYFKTSVAANPFEVMLAQVMDDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPF
AQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKEL
LSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFIS
SGSFGATINAWQDDWNARAWNTSFATNKSQFLWGDAQVFNQGYGLLGFKNAKHYYQVDG
KRGKGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKM
GSKVTFKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQ
INMVLNRNTKIVEDMEKVKATERFLPTHPTGLLKTGTIDRHQKTFDSQTDISFGKDFIEV
IPWQLNFSDFSSQKIHDYFKHYGVKELEIESIALGLGANSKENTLIKMDYRLKNWER
PDTKTFKDSYVLRK

>SEQ ID NO 2852:62_2603 frame: 3

LKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFLHLSN
MGANTVRVVKPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLK
REAGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKKQY
KGRYFKTSVAANPFEVMLAQVMDDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPFQAQ
PKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLQ
GYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFISGS
FGATINAWQDDWNARAWNTSFATNKSQFLWGDAQVFNQGYGLLGFKNAKHYYQVDGKRG
KGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMNGSK
VTFKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINM
VLRNTKIVEDMEKVKATERFLPTHPTGLLKTGTIDRHQKTFDSQTDISFGKDFIEVRIPW
QLNFSDFSSQKIHDYFKHYGVKELEIESIALGLGANSKENTLIKMDYRLKNWERPDT
KTFKDSYYSIKKEWSKERERTYGP

**Table 28: C mparative Sequences relating to SAG1552
(conserved hypothetical protein)**

>SEQ ID NO 2853:62_A909 frame: 1

KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFHL
ISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRG
YLKREAGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKK
TQYKGRYFKTSVAANPFEVMLAQVMDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPF
AQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELS
LSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFIS
SGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFKNAKHHYQVDG
KRGKGWKKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMN
GSKVTFSSKSSDFVLSIDPNGKSELFQERYNALKANYLRQLNGKDFYAFPPKKNSSNFQ
INMVLNRTKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFDSQTDISFGKDFIEVR
IPWQLLNFSQSSQRIHDDYFKHYGVKELEN.EPLL.D.VLIAKKTTH..RWQIIV.KIGR
DPIPKPF.K

>SEQ ID NO 2854:62_A909 frame: 1

KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFHL
ISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRG
YLKREAGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKK
TQYKGRYFKTSVAANPFEVMLAQVMDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPF
AQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELS
LSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFIS
SGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFKNAKHHYQVDG
KRGKGWKKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMN
GSKVTFSSKSSDFVLSIDPNGKSELFQERYNALKANYLRQLNGKDFYAFPPKKNSSNFQ
INMVLNRTKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFDSQTDISFGKDFIEVR
IPWQLLNFSQSSQRIHDDYFKHYGVKELEN.EPLL.D.VLIAKKTTH..RWQIIV.KIGR
DPIPKPF.K

>SEQ ID NO 2855:62_CJB110 frame: 1

YYFDGSLYLPKGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPIT
QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
ITAFNDNYRGYLKREAGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGT
VAYTNHQEKKTKYKGRYFKTSVAANPFEVMLAQVMDELTHYETAKYGWQHLISFSNSPTT
DPFHYRKPFQAPAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
EDRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQ
LLEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHNQFLWGDAQVFNQGYGLLGFK
NAKHHYQVDGKRGKGWKKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
TPKSGSRKMNGSKVTFSSKSSDFVLSIDPNGKSELFQERYNALKANYLRQLNGKDFYAF
PKKNSSNFQINMVLNRTKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFDSQTDI
SFGKDFIEVRIPWQLLNFSQSSQRIHDDYFKHYGVKELEIESIALGLGANSKENTLIK
ADYRLKNWERPDTKTFLKDSYYVLRK

>SEQ ID NO 2856:62_COH1 frame: 2

LPQGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWF
HLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNY
RGLYKREAGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQE
KKTQYKGRYFKTSVAANPFEVMLAQVMDELTHYETAKYGWQHLISFSNSPTTDPFHYRK
FEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKE
LSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESF
ISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFKNAKHHYQV
DGKRGKGWKKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRK
MNGSKVTFSSKSSDFVLSIDPNGKSELFQERYNALKANYLRQLNGKDFYAFPPKKNSSNF
EQINMVLNRTKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFDSQDPDISFGKDFIE
VRIPWQLLNFSQSSQRIHDDYFKHYGVKELEIESIALGLGANSKENTLIKMAKYRLKNW
ERPDKTFLKD

>SEQ ID NO 2857:62_H36B frame: 2

RGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFHL
ISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRG
YLKREAGVVDILHGRKQVWNTDFGSSHYHYDLSFWVLGYVVGDDGHSQVVALY

>SEQ ID NO 2858:62_JM9130013 frame: 3

FVVKGDTVLHKPTNKPFFVVKGVDESSLAGYHHNDFPITQKTYREWFHLISNMGANTVRV
KVPNMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLKREAGVVD
ILHGRKQVWNTDFGSSHYHYDLSFWVLGYVVGDDWNSGTVAYTNHQEKKTKYKGRYFKTS
VAANPFEVMLAQVMDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPFQAPAPKYVQLNV
ENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLSQGYVKLLNA
YHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFISSGSFGATINAW

**Table 28: Comparative Sequences relating to SAG1552
(conserved hypothetical protein)**

QDDWNARVWNTSFATNKHQSFLWGDAQVFNQGYGLLGFKNAKHHYQVDGKRGKEWKHPL
MTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMNGSKVTFSKSSD
FVLSIDPNGKSELFVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINMVLNRTKIV
EDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFSQTDISFGKDFIEVRIPWQLNFSDF
SSQKIHDDYFKHYGVKELEIESIALGLGANSKENTLIKMDYRLKNWERPDTKTFLKDSY
YSIKK

>SEQ ID NO 2859:62_M732 frame: 2

TRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPITQKTYREWFHLISNMGAN
TVRVKVP MNVAFYDALYHHNKESKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLKREAK
GVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDCNSGTVAYTNHQEKKTYKGRY
FKTSVAANPFEVMLAQVMDDELTHYETAKYGWQHLISFSNSPTTDPFHYRKPFEAQAPKYV
QLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLSQGYVK
LLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRLLEDYESFISSGSFGAT
INAWQDDWNARAWNTSFATNKHQSFLWGDAQVFNQGYGLLGFKNAKHHYQVDGKRGKGEW
KHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITPKSGSRKMNGSKVTFS
KSSDFVLSIDPNGKSELFVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINMVLN
TKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFSQTDISFGKDFIEVRIPWQLN
FSDPSSQKIHDDYFKHYGVKELEIESIALGLGANSKENTLIKMDYRLKNWERPDTKTFL
KDSYYSIK

>SEQ ID NO 2860:62_M781 frame: 1

FDGSLYLPQGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPITQK
TYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKESKRPLYLLQGIRIDSYRNNASIT
AFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGTVA
YTNHQEKKTYKGRYFKTSVAANPFEVMLAQVMDDELTHYETAKYGWQHLISFSNSPTTDP
FHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISKED
RQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQRL
EDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHQSFLWGDAQVFNQGYGLLGFKNA
KHHYQVDGKRGKEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDITP
KSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELFVQERYNALKANYLRQLNGKDFYAFPPK
KNSSNFEQINMVLNRTKIVEDMEKVKATERFLPTHPTGLLKTGTTDRHQKTFSQTDISF
GKDFIEVRIPWQLNFSDFSSQKIHDDYFKHYGVKELEIESIALGLGANSKENTLIKMD
YRLKNWERPDTKTFLKDSYYSIKKEW

SEQ2850	-----FVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2851	-----KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2852	-----LKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2853	-----KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2854	-----KGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2855	YFDGSLYLPKGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2856	-----LPQGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2857	-----RGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2858	-----FVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2859	-----TRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT
SEQ2860	-FDGSLYLPQGLLKENTRTNFVVKGDTVLHKPTNKPFFVVKGV DVESSLAGYHHNDFPIT

SEQ2850	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2851	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2852	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2853	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2854	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2855	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2856	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2857	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2858	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2859	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS
SEQ2860	QKTYREWFHLISNMGANTVRVKVPMNVAFYDALYHHNKASKRPLYLLQGIRIDSYRNNAS

SEQ2850	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2851	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2852	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2853	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2854	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2855	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2856	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2857	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSSSHYHYDLSFWVLGYVVGDDGHSST
SEQ2858	ITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDFGSSSHYHYDLSFWVLGYVVGDDWNSGT

SEQ2859	ITAFNDNYRGYLKREAGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDCNSGT
SEQ2860	ITAFNDNYRGYLKREAGVVDILHGRKQVWNTDFGSRHYHYDLSFWVLGYVVGDDWNSGT
SEQ2850	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2851	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2852	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2853	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2854	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2855	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2856	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2857	VALY-----
SEQ2858	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2859	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2860	VAYTNHQEKKTQYKGRYFKTSAANPFEVMLAQVMDELTHYETAKYGWQHLSFSNSPTT
SEQ2850	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGIFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2851	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2852	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2853	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2854	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2855	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2856	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2857	-----
SEQ2858	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2859	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2860	PFHYRKPFEAQAPKYVQLNVENIQANSNVKAGMFAAYKAIDFHPRYKDYLLFDKENISK
SEQ2850	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2851	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2852	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2853	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2854	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2855	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2856	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2857	-----
SEQ2858	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2859	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2860	DRQKIKELSLSQGYVKLLNAYHKIPVLVTGYGYSTARGIAQKEIDKRPLPINEKEQGQR
SEQ2850	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2851	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2852	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2853	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2854	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2855	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2856	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2857	-----
SEQ2858	LEDYESFISSGSFGATINAWQDDWNARVWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2859	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2860	LEDYESFISSGSFGATINAWQDDWNARAWNTSFATNKHSQFLWGDAQVFNQGYGLLGFK
SEQ2850	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2851	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2852	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2853	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2854	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2855	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2856	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2857	-----
SEQ2858	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2859	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2860	AKHHYQVDGKRKGGEWKHPLMTSATGDDLYASSDESILYLAIKTKPEKLKEKRLLPIDI
SEQ2850	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP
SEQ2851	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP
SEQ2852	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP
SEQ2853	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP
SEQ2854	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP
SEQ2855	PKSGSRKMNGSKVTFSKSSDFVLSIDPNGKSELVQERYNALKANYLRQLNGKDFYAFP

၇၅၃၇၅၃

16

Table 29: Comparative Sequences relating to SAG1641 (YaeC family protein)

SEQ ID NO. 2901: SAG1641 FROM THE 090 GBS TYPE Ia STRAIN

AATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAA
GCACGTTGGGATAAAATTGAAAAGCTAGTAGGCGATAAAGCTAAAATCAAATTCACAGAATTTACAGATTATACACAA
CCAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAG
GAAAAATAAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCCCCAATTCGTATCTATTCTGAGAAGGTAAAATCT
CTTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
CAGTCAGCAGGTTTAATCAAATTGAATGTTTCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAA
GATATTAATATTCAGGAGTTAGATGCGAGTCAAACACCACGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAAT
ACATACATTGAGCAAGCTAATTTAAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAAATTCAAAACAATGG
ATTAATATCATTGCGGGACGTAAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTAT
CACACAGATGAAGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGGAACCCAGCTTTCTTGTACAA

SEQ ID NO. 2902: SAG1641 FROM THE 1169NT1 GBS TYPE V STRAIN

(REVERSE COMPLEMENT)

ATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAAG
CAGCTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAAC
CAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAGG
AAAATAAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCTC
TTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
AGTCAGCAGGTTTAATCAAATTGAATGTTTCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAG
ATATTAATATTCAGGAGTTAGATGCGAGTCAAACACCACGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAATA
CATACATTGAGCAAGCTAATTTAAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAAATTCAAAACAATGGA
TTAATATCATTGCGGGACGTAAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATC
ACACAGATGAAGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

SEQ ID NO. 2903: SAG1641 FROM THE 18RS21 GBS TYPE II STRAIN

AATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAA
GCACGTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAA
CCAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAG
GAAAAATAAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCT
CTTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
CAGTCAGCAGGTTTAATCAAATTGAATGTTTCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAG
GATATTAATATTCAGGAGTTAGATGCGAGTCAAACACCACGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAAT
ACATACATTGAGCAAGCTAATTTAAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAAATTCAAAACAATGG
ATTAATATCATTGCGGGACGTAAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTAT
CACACAGATGAAGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCAC

SEQ ID NO. 2904: SAG1641 FROM THE 2603 V/R GBS TYPE V STRAIN

AATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAA
GCACGTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAA
CCAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAG
GAAAAATAAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCT
CTTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
CAGTCAGCAGGTTTAATCAAATTGAATGTTTCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAG
GATATTAATATTCAGGAGTTAGATGCCAGTCAAACACCACGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAAT
ACATACATTGAGCAAGCTAATTTAAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAAATTCAAAACAATGG
ATTAATATCATTGCGGGACGTAAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTAT
CACACAGATGAAGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

SEQ ID NO. 2905: SAG1641 FROM THE A909 GBS TYPE Ia STRAIN

AATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAA
GCACGTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAA
CCAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAG
GAAAAATAAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCT
CTTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
CAGTCAGCAGGTTTAATCAAATTGAATGTTTCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAG
GATATTAATATTCAGGAGTTAGATGCGAGTCAAACACCACGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAAT
ACATACATTGAGCAAGCTAATTTAAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAAATTCAAAACAATGG
ATTAATATCATTGCGGGACGTAAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTAT
CACACAGATGAAGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

Table 29: Comparative Sequences relating to SAG1641 (YaeC family protein)

SEQ ID NO. 2906: SAG1641 FROM THE CJB110 GBS NONTYPEABLE STRAIN

AAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAAGCACGTTGGGATAAAATTGAAAAGCTAGT
AGGCGATAAAGCTAAAATCAAATTCACAGAATTTACAGATTATACACAACCAAATCAAGCGACAGCCAATAAGGATGT
GGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAGGAAAATAAGAAAACTTAATTCACCTTGA
AAAGACTTACTTAGCCCCAATTCGTATCTATTCTGAGAAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTAT
TGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGT
TTCTGGTAAGAAGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGATATTAATATTAGGAGTTAGATGCGAG
TCAAACACCCAGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAATACATACATTGAGCAAGCTAATTTAAACC
TTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAATTCAAACAATGGATTAATATCATTGCGGGACGTAAAAATTG
GAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAAAAGTTATCAA
AGATACTTCAGCTGATATTCCACAATGGAA

**SEQ ID NO. 2907: SAG1641 FROM THE COH1 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)**

AGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAAGCACGTTG
GGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAACCAAATCA
AGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAGGAAAATAA
GAAAACTTAATTCACCTTGAAGAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCTCTTAAAA
ATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTACTTCAGTCAGC
AGGTTTAAATCAAATTGAATGTTTCTGGTAAGAAGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAA
TATTAGGAGTTAGATGCGAGTCAAACACCCAGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAATACATACAT
TGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAATTCAAACAATGGATTAATAT
CATTGCGGGACGTAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGA
TGAAGTGAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

SEQ ID NO. 2908: SAG1641 FROM THE H36b GBS TYPE Ib STRAIN

AAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAAGCAC
GTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAACCAA
ATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAGGAAA
ATAAGAAAACTTAATTCACCTTGAAGAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCTCTTA
AAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTTCAGT
CAGCAGGTTTAAATCAAATTGAATGTTTCTGGTAAGAAGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATA
TTAATATTAGGAGTTAGATGCGAGTCAAACACCCAGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAATACAT
ACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAATTCAAACAATGGATTA
ATATCATTGCGGGACGTAAAAATTGAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACA
CAGATGAAGTGAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

SEQ ID NO. 2909: SAG1641 FROM THE JM3190013 GBS TYPE VIII STRAIN

TTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAAGCACGTTGGGA
TAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAACCAAATCAAGC
GACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAGGAAAATAAGAA
AACTTAATTCACCTTGAAGAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCTCTTAAAAAATT
GAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTTCAGTCAGCAGG
TTTAAATCAAATTGAATGTTTCTGGTAAGAAGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATAT
TCAGGAGTTAGATGCGAGTCAAACACCCAGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAATACATACATTGA
GCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAATTCAAACAATGGATTAATATCAT
TGCGGGACGTAAAAATTGAAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGA
AGTGAAAAAAGTTATCAAAGATACTTCAGCTGATATTCCACAATGG

SEQ ID NO. 2910: SAG1641 FROM THE M732 GBS TYPE III STRAIN

AATCAAGAAGTTTCAGCAAGCTCAACTTCAAGTAAAGTTGTTAAAGTTGGTGTATGACCTTTTCTGACACTGAAAAA
GCACGTTGGGATAAAATTGAAAAGCTAGTAGGTGATAAAGCTAAAATCAAATTTACAGAATTTACAGATTATACACAA
CCAAATCAAGCGACAGCCAATAAGGATGTGGATATTAATGCCTTTCAACATTACAATTTCTTAGAAAACCTGGAATAAG
GAAAAAAGGAAAACTTAATTCACCTTGAAGAGACTTACTTAGCTCCAATTCGTATCTATTCTGAGAAGGTAAAATCT
CTTAAAAAATTGAAAAAAGGAGCCACTATTGCAATTCCAAATGATGCAACAAATGGTAGCCGTGCATTGTATGTCCTT
CAGTCAGCAGGTTTAAATCAAATTGAATGTTTCTGGTAAGAAGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAG
GATATTAATATTAGGAGTTAGATGCGAGTCAAACACCCAGTGCACCTCAAAGATGTAGATGCAGCTATTATTAATAAT
ACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAATCAGATAAAAATTCAAACAATGG
ATTAATATCATTGCGGGACGTAAAAATTGAAAAAAGCAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTAT
CACACAGATGAAGTGAAAAAAGTTATCAAAGATAC

0011- SAC1641 FROM THE M781 GBS TYPE III STRAIN

SEQ2901 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC CCCAATT CGTATCT ATTCTGAG
 SEQ2902 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2903 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2904 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2905 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2906 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2907 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG
 SEQ2908 AAGAAAAA CTTAATTC CACTTGAA AAGACTT ACTTAGC TCCAATT CGTATCT ATTCTGAG

Table 29: Comparative Sequences relating to SAG1641 (YaeC family protein)

SEQ2909	AAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAG
SEQ2910	AAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAG
SEQ2911	AAGAAAACTTAATTCACCTTGAAAAGACTTACTTAGCTCCAATTCGTATCTATTCTGAG
SEQ2901	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2902	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2903	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2904	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2905	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2906	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2907	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2908	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2909	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2910	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2911	AAGGTAAAATCTCTTAAAAAATTGAAAAAGGAGCCACTATTGCAATTCCAAATGATGCA
SEQ2901	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2902	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2903	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2904	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2905	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2906	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2907	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2908	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2909	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2910	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2911	ACAAATGGTAGCCGTCATTGTATGTCCTTCAGTCAGCAGGTTTAATCAAATTGAATGTT
SEQ2901	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2902	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2903	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2904	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2905	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2906	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2907	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2908	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2909	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2910	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2911	TCTGGTAAGAAGGTTGCAACAGTTGCTAATATCACATCTAATAAAAAAGGATATTAATATT
SEQ2901	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2902	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2903	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2904	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2905	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2906	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2907	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2908	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2909	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2910	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2911	CAGGAGTTAGATGCGAGTCAAACACCACGTGCACTCAAAGATGTAGATGCAGCTATTATT
SEQ2901	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2902	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2903	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2904	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2905	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2906	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2907	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2908	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2909	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2910	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA
SEQ2911	AATAATACATACATTGAGCAAGCTAATTTAAACCTTCAGATGCTATCTTTGTTGAGAAA

Table 29: Comparative Sequences relating to SAG1641 (YaeC family protein)

SEQ2901	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2902	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2903	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2904	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2905	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2906	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2907	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2908	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2909	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2910	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG
SEQ2911	TCAGATAAAAAATTCAAAACAATGGATTAATATCATTGCGGGACGTAAAAATTGGAAAAAG

SEQ2901	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2902	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2903	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2904	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2905	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2906	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2907	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2908	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2909	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2910	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA
SEQ2911	CAAAAGAACGCTAAAGCTATCCAAGCTATCTTGGATGCTTATCACACAGATGAAGTGAAA

SEQ2901	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGGAACCCAGCTTTCTTGTACAA
SEQ2902	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2903	AAAGTTATCAAAGATACTTCAGCTGATATCCAC-----
SEQ2904	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2905	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2906	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGGAA-----
SEQ2907	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2908	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2909	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----
SEQ2910	AAAGTTATCAAAGATAC-----
SEQ2911	AAAGTTATCAAAGATACTTCAGCTGATATCCACAATGG-----

>SEQ ID NO 2950: 35_090 frame: 1
 NQEVSSSTSSKVVKGVMFTSDTEKARWDKIEKLVGDKAKIKFTEFTDYTQPNQATANK
 DVDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
 TNGSRALYVLQSAGLIKLVNVSCKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
 NNTYIEQANLKPDAIFVEKSDKNSKQWINIIAGRNWKKQKNAKAIQAILDAYHTDEVK
 VKIKDTSADIPQWNPFLY

>SEQ ID NO 2951: 35_1169NT frame: 3
 QEVSSSTSSKVVKGVMFTSDTEKARWDKIEKLVGDKAKIKFTEFTDYTQPNQATANKD
 VDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDAT
 NGSRALYVLQSAGLIKLVNVSCKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIIN
 NNTYIEQANLKPDAIFVEKSDKNSKQWINIIAGRNWKKQKNAKAIQAILDAYHTDEVKK
 VIKDTSADIPQW

>SEQ ID NO 2952: 35_18RS21 frame: 1
 NQEVSSSTSSKVVKGVMFTSDTEKARWDKIEKLVGDKAKIKFTEFTDYTQPNQATANK
 DVDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
 TNGSRALYVLQSAGLIKLVNVSCKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
 NNTYIEQANLKPDAIFVEKSDKNSKQWINIIAGRNWKKQKNAKAIQAILDAYHTDEVK
 VKIKDTSADIP

>SEQ ID NO 2953: 35_2603 frame: 1
 NQEVSSSTSSKVVKGVMFTSDTEKARWDKIEKLVGDKAKIKFTEFTDYTQPNQATANK
 DVDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
 TNGSRALYVLQSAGLIKLVNVSCKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
 NNTYIEQANLKPDAIFVEKSDKNSKQWINIIAGRNWKKQKNAKAIQAILDAYHTDEVK
 VKIKDTSADIPQW

Table 29: C mparativ Sequences relating to SAG1641 (YaeC family protein)

>SEQ ID NO 2954:35_A909 frame: 1
 NQEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
 DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
 TNGSRALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAI
 NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
 KVIKDTSDIPQW

>SEQ ID NO 2955:35_CJB110 frame: 2
 SKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANKDVDINAFQHY
 NFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATNGSRALYVL
 QSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIINNTYIEQANL
 KPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVKKVIKDTSDI
 PQW

>SEQ ID NO 2956:35_COH1 frame: 2
 VSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANKDVD
 INAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATNG
 SRALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIINNT
 YIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVKKVI
 KDTSDIPQW

>SEQ ID NO 2957:35_H36B frame: 3
 EVSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANKDV
 DINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATN
 GSRALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIINN
 TYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVKKV
 IKDTSDIPQW

>SEQ ID NO 2958:35_JM9130013 frame: 2
 SASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANKDVDI
 NAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATNGS
 RALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIINNTY
 IEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVKKVIK
 DTSADIPQW

>SEQ ID NO 2959:35_M732 frame: 1
 NQEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
 DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
 TNGSRALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAI
 NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
 KVIK

>SEQ ID NO 2960:35_M781 frame: 2
 VSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANKDVD
 INAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATNG
 SRALYVLQSAGLIKLVSGKKVATVANITSNKKDINIQLDASQTPRALKDVDAAIINNT
 YIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAIWDAYHTDEVKKVI
 KDTSDIPQW

SEQ2950	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2951	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2952	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2953	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2954	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2955	-----SKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2956	--VSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2957	-EVSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2958	---SASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2959	QEVSSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK
SEQ2960	--VSASSTSSKVVKGVMFTFSDTEKARWDKIEKLVGDKAKIKFTEFTDYTPNQATANK

SEQ2950	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2951	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2952	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2953	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2954	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2955	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2956	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2957	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2958	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA

Table 29: Comparative Sequences relating to SAG1641 (YaeC family protein)

SEQ2959	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2960	DVDINAFQHYNFLENWNKENKKNLIPEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDA
SEQ2950	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2951	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2952	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2953	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2954	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2955	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2956	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2957	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2958	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2959	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2960	TNGSRALYVLQSAGLIKLNVS GKKVATVANITSNKKDINIQLDASQTPRALKDVDAAII
SEQ2950	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2951	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2952	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2953	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2954	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2955	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2956	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2957	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2958	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2959	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2960	NNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRKNWKKQKNAKAIQAILDAYHTDEVK
SEQ2950	KVIKDT SADIPQWNPAPFLY
SEQ2951	KVIKDT SADIPQW-----
SEQ2952	KVIKDT SADIP-----
SEQ2953	KVIKDT SADIPQW-----
SEQ2954	KVIKDT SADIPQW-----
SEQ2955	KVIKDT SADIPQW-----
SEQ2956	KVIKDT SADIPQW-----
SEQ2957	KVIKDT SADIPQW-----
SEQ2958	KVIKDT SADIPQW-----
SEQ2959	KVIKD-----
SEQ2960	KVIKDT SADIPQW-----

**Table 30: Comparative Sequences relating to SAG2147
(protein f unknown function / lipoprotein, putative)**

**SEQ ID NO. 3001: SAG2147 FROM THE 1169NT1 GBS TYPE V STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAAAGCAGATAAAAGTTCGCGTAGCC
AAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGTAGAAGATGTAAAACAGGCT
CCAAAACCTTCTCAGGCATCTAATGAAGTCCCAAATCAAGTTCTCAATCTACAGAAGCT
AATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACA
GAAAATACCCCTGCTACCAGTCAGGCACAACAACTTATGCTGTTACTGAGACAACCTTAC
AAACCTGCTCAACACCAGACAAGTGGCCAAGTATTGAGCAATGGAAATACTGCAGGGGCG
GTCGGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTTGG
GAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAGCT
TCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAGTT
AATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC

**SEQ ID NO. 3002: SAG2147 FROM THE 18RS21 GBS TYPE II STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAAAGCAGATAAAAGTTC
GCGTAGCCAAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGTAGAAGATGTAA
AACAGGCTCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTA
CAGAAGCTAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCAGCTGTAGAACAAGCAG
TTGTAACAGAAAACACCCCTGCTACCAGTCAGGCACAACAAGCTTATGCTGTTACTGAGA
CAACTTATAGACCTGCTCAACACCAGACGAGTGGCCAAGTATTGAGTAATGGAAATACTG
CAGGGGCTATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGT
CTACTTGGGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCT
CAGGAGCTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGG
ATCAAGTTAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC

**SEQ ID NO. 3003: SAG2147 FROM THE 2603 V/R GBS TYPE V STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAAAGCAGATAAAAGT
TCGCGTAGCCAAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGTAGAAGATGT
AAAACAGGCTCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATC
TACAGAAGCTAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCAGCTGTAGAACAAGC
AGTTGTAACAGAAAACACCCCTGCTACCAGTCAGGCACAACAAGCTTATGCTGTTACTGA
GACAACTTATAGACCTGCTCAACACCAGACGAGTGGCCAAGTATTGAGTAATGGAAATAC
TGCAGGGGCTATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCA
GTCTACTTGGGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGC
CTCAGGAGCTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCA
GGATCAAGTTAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTA
C

**SEQ ID NO. 3004: SAG2147 FROM THE 090 GBS TYPE Ia STRAIN
(REVERSE COMPLEMENT)**

TAGCCAAAAAATCAAAAATGATTAAGGCGACATCTAAATCAAAAGTAGAAGATGTAAAAC
AGGCTCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAG
AAGCTAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCAGCTGTAGAACAAGCAGTTG
TAACAGAAAACACCCCTGCTACCAGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAA
CTTATAGACCTGCTCAACACCAGACGAGTGGCCAAGTATTGAGTAATGGAAATACTGCAG
GGGCTATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTA
CTTGGGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAG
GAGCTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGA

**SEQ ID NO. 3005: SAG2147 FROM THE A909 GBS TYPE Ia STRAIN
(REVERSE COMPLEMENT)**

AAGGCGACATCTAAATCAAAAGTAGAAGATGTAAAACAGGCTCCAAAACCTTCTCAGGCA
TCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAATTCTCAGCAACAAGTT
ACTGCGAGTGAAGAGGCAGCTGTAGAACAAGCAGTTGTAACAGAAAACACCCCTGCTACC
AGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTATAGACCTGCTCAACACCAG
ACAAGTGGCCAAGTATTGAGTAATGGAAATACTGCAGGGGCTATTGGCTCAGCAGCTGCA
GCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTTGGGAACATATTATTGCCCGT
GAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAGCTTCAGGACTTTTCCAAACG
ATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGAATCAAGTTAATTCAGCTATTAAAGCT
TATCGTGCTCAAGGTTTATCA

**Table 30: Comparative Sequences relating to SAG2147
(protein of unknown function / lipoprotein, putative)**

**SEQ ID NO. 3006: SAG2147 FROM THE CJB110 GBS NONTYPEABLE STRAIN
(REVERSE COMPLEMENT)**

AATCTTTGTCAAAGCAGATAAAGTTCGCGTAGCCAAAAATCAAAAATGACTAAGGCGA
CATCTAAATCAAAAGTAGAAGATGTAAAACAGGCTCCAAAACCTTCTCAGGCATCTAATG
AAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAATTCTCAGCAACAAGTTACTGCGA
GTGAAGAGGCAGCTGTAGAACAAGCAGTTGTAACAGAAAACACCCCTGCTACCAGTCAGG
CACAACAAGCTTATGCTGTTACTGAGACAACCTTATAGACCTGCTCAACACCAGACGAGTG
GCCAAGTATTGAGTAATGGAAATACTGCAGGGGCTATTGGCTCAGCAGCTGCAGCACAAA
TGGCTGCTGCAACAGGAGTCCCTCAGTCTACTTGGGAACATATTATTGCCCGTGAATCAA
ATGGTAATCCTAATGTTGCTAATGCCTCAGGAGCTTCAGGACTTTTCCAAACGATGCCAG
GTTGGGGTTCAACAGCTACAGTTCAAGTCAAGTTAATTACAGCTATTAAAGCTTATCGTG
CTCAAGGTTTATCAGCTTGGGGTTAC

**SEQ ID NO. 3007: SAG2147 FROM THE COH1 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAA
AGTTCGCGTAGCCAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGTAGAAGA
TGTAACACAGGCTCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCA
ATCTACAGAAGCTAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACA
AGCAGTTGTAACAGAAAATACCCCTGCTACCAGTCAGGCACAACAACTTATGCTGTTAC
TGAGACAACCTTACAAACCTGCTCAACACCAGACAAGTGCCCAAGTATTGAGCAATGGAAA
TACTGCAGGGGCGGTCGGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCC
TCAGTCTACTTGGGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAA
TGCCTCAGGAGCTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGT
TCAGGATCAAGTTAATTACAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGG
TTAC

**SEQ ID NO. 3008: SAG2147 FROM THE H36b GBS TYPE Ib STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGC
AGATAAAGTTTCGCGTAGCCAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGT
AGAAGATGTAAACAGGCTCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAG
TTCTCAATCTACAGAAGCTAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCAGCTGT
AGAACAAGCAGTTGTAACAGAAAACACCCCTGCTACCAGTCAGGCACAACAAGCTTATGC
TGTTACTGAGACAACCTTATAGACCTGCTCAACACCAGACAAGTGGCCAAGTATTGAGTAA
TGGAAATACTGCAGGGGCTATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGG
AGTCCCTCAGTCTACTTGGGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGT
TGCTAATGCCTCAGGAGCTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGC
TACAGTTCAGGATCAAGTTAATTACAGCTATTAAAGCTT

**SEQ ID NO. 3009: SAG2147 FROM THE M732 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)**

AAAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAGTTCGCGTAGC
CAAAAATCAAAAATGACTAAGGCGACATCTAAATCAAAAGTAGAAGATGTAAACAGGC
TCCAAAACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGC
TAATTCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAAC
AGAAAATACCCCTGCTACCAGTCAGGCACAACAACTTATGCTGTTACTGAGACAACCTTA
CAAACCTGCTCAACACCAGACAAGTGGCCAAGTATTGAGCAATGGAAATACTGCAGGGGC
GGTCCGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTTG
GGAACATATTATTGCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAGC
TTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAGT
TAATTACAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTA

**Table 30: Comparativ Sequences relating to SAG2147
(prot in of unknown function / lipoprotein, putative)**

SEQ ID NO. 3010: SAG2147 FROM THE M781 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)

GTAACCCCAAGCTGATAAACCTTGAGCACGATAAGCTTTAATAGCTGAATTAACCTTGATC
CTGAACGTAGCTGTTGAACCCCAACCTGGCATCGTTTGGAAAAGTCCTGAAGCTCCTGA
GGCATTAGCAACATTAGGATTACCATTTGATTCACGGGCAATAATATGTTCCCAAGTAGA
CTGAGGGACTCCTGTTGCAGCAGCCATTTGTGCTGCAGCAGCAGATCCGACCGCCCCCTGC
AGTATTTCCATTGCTCAATACTTGGCCACTTGTCTGGTGTGAGCAGGTTTGTAAAGTTGT
CTCAGTAACAGCATAAGTTTGTGTGCCTGACTGGTAGCAGGGGTATTTTCTGTTACAAC
TGCTTGTCTACAGCCGCCTCTTCACTCGCAGTAACCTTGTGCTGAGAATTAGCTTCTGT
AGATTGAGAACTTGATTTTGGGGCTTCATTAGATGCCTGAGAAGGTTTTGGAGCCTGTTT
TACATCTTCTACTTTTGATTTAGATGTCGCCTTAGTCATTTTTTGATTTTTTGGCTACGCG
AACTTTATCTGCTTTTGACAAAGA

SEQ3001 -----
SEQ3002 -----
SEQ3003 -----
SEQ3004 -----
SEQ3005 AGGCGACATCTAAATCAAAAGTAGAAGATGTAAACAGGCTCCAAAACCTTCTCAGGCA
SEQ3007 -----
SEQ3008 -----
SEQ3009 -----
SEQ3010 -----

SEQ3001 -----
SEQ3002 -----
SEQ3003 -----
SEQ3004 -----
SEQ3005 CTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAATTCTCAGCAACAAGTT
SEQ3007 -----
SEQ3008 -----
SEQ3009 -----
SEQ3010 -----

SEQ3001 -----
SEQ3002 -----
SEQ3003 -----
SEQ3004 -----
SEQ3005 CTGCGAGTGAAGAGGCAGCTGTAGAACAAGCAGTTGTAACAGAAAACACCCCTGCTACC
SEQ3007 -----
SEQ3008 -----
SEQ3009 -----
SEQ3010 -----

SEQ3001 -----
SEQ3002 -----
SEQ3003 -----
SEQ3004 -----
SEQ3005 GTCAGGCACAACAAGCTTATGCTGTACTGAGACAACTTATAGACCTGCTCAACACCAG
SEQ3007 -----
SEQ3008 -----
SEQ3009 -----
SEQ3010 -----

SEQ3001 -----
SEQ3002 -----
SEQ3003 -----
SEQ3004 -----
SEQ3005 CAAGTGGCCAAGTATTGAGTAATGGAAATACTGCAGGGGCTATTGGCTCAGCAGCTGCA
SEQ3007 -----
SEQ3008 -----
SEQ3009 -----
SEQ3010 -----

**Table 30: Comparative Sequences relating to SAG2147
(protein f unknown function / lip protein, putative)**

SEQ3001	-----
SEQ3002	-----
SEQ3003	-----
SEQ3004	-----
SEQ3005	CACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTTGGGAACATATTATTGCCCGT
SEQ3007	-----
SEQ3008	-----
SEQ3009	-----
SEQ3010	-----
SEQ3001	-----
SEQ3002	-----
SEQ3003	-----
SEQ3004	-----
SEQ3005	AATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAGCTTCAGGACTTTTCCAAACG
SEQ3007	-----
SEQ3008	-----
SEQ3009	-----
SEQ3010	-----
SEQ3001	-----
SEQ3002	-----
SEQ3003	-----
SEQ3004	-----
SEQ3005	TGCCAGGTTGGGGTTCAACAGCTACAGTTCAGAATCAAGTTAATTAGCTATTAAAGCT
SEQ3007	-----
SEQ3008	-----
SEQ3009	-----
SEQ3010	-----
SEQ3001	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3002	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3003	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3004	-----TAGCCAAA
SEQ3005	ATCGTGCTCAAGGTTTATCASAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3007	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3008	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3009	AAAGTTCACAAGTTACTACTGAATCTTTGTCAAAGCAGATAAAAGTTCGCGTAGCCAAA
SEQ3010	-----GTAACCCCAAGCTGA-----TAAACCTTGAGCAGGATAAGCTTTAATAGCTGAA
SEQ3001	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3002	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3003	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3004	AATCAAAAATGATTAAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3005	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3007	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3008	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3009	AATCAAAAATGACTAAGGCGACATCTAAATCAAAGTAGAAGATGTAAACAGGCTCCA
SEQ3010	TAACCTTGATCCTGAACCTGTAGCTGTTGAACCCCAACCTGGCATCGTTTGGAAAAGTCTT
SEQ3001	AACCTTCTCAGGCATCTAATGAAGTCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3002	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3003	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3004	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3005	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3007	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3008	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3009	AACCTTCTCAGGCATCTAATGAAGCCCCAAAATCAAGTTCTCAATCTACAGAAGCTAAT
SEQ3010	AAGCTCCTGAGGCATT---AGCAACATTAGGATTAC-CATTTGATTCACGGGCAATAAT
SEQ3001	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3002	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3003	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3004	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3005	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3007	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3008	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3009	TCTCAGCAACAAGTTACTGCGAGTGAAGAGGCGGCTGTAGAACAAGCAGTTGTAACAGA
SEQ3010	TGTTCCCAAGTAGACTGAGGACTCCTGTTGCAGCAGCCATTTGTGCTGCAGCAGCAGA

**Table 30: Comparative Sequences relating to SAG2147
(protein of unknown function / lip protein, putative)**

SEQ3001	--AAATACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3002	--AAACACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3003	--AAACACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3004	--AAACACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3005	--AAACACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3007	--AAATACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3008	--AAACACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3009	--AAATACCCCTGCTACCACTCAGGTCAGGCACAACAAGCTTATGCTGTTACTGAGACAACCTTA
SEQ3010	CCGACCGCCCTGCAGTATTTCCATTGCTCAATACTTG-GCCACTTGTCTGGTGTGAG
SEQ3001	AAACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGCAATGGAAATACTGCAGGGG
SEQ3002	AGACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGTAATGGAAATACTGCAGGGG
SEQ3003	AGACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGTAATGGAAATACTGCAGGGG
SEQ3004	AGACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGTAATGGAAATACTGCAGGGG
SEQ3005	AGACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGTAATGGAAATACTGCAGGGG
SEQ3007	AAACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGCAATGGAAATACTGCAGGGG
SEQ3008	AGACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGTAATGGAAATACTGCAGGGG
SEQ3009	AAACCTGCTCAACACCAGACAGTGGC-CAAGTATTGAGCAATGGAAATACTGCAGGGG
SEQ3010	AGGTTTGTAAAGTTGTCTCAGTAACAGCATAAGTTTGTGTGCCTGACTGGTAGCAGGGG
SEQ3001	GGTCGGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3002	TATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3003	TATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3004	TATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3005	TATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3007	GGTCGGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3008	TATTGGCTCAGCAGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3009	GGTCGGATCTGCTGCTGCAGCACAAATGGCTGCTGCAACAGGAGTCCCTCAGTCTACTT
SEQ3010	A-TTT--TCTGTTACAACCTGCTTGTCTACAGCCGCTCTTCACTCGCAGTAACCTGTT
SEQ3001	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3002	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3003	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3004	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3005	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3007	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3008	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3009	GGGAACATATTATTGCCCCGTGAATCAAATGGTAATCCTAATGTTGCTAATGCCTCAGGAG
SEQ3010	GCTGAGA-ATTAGCTTCTGTAGATTGAG---AA--CTTGATTTTGGGGCTTCATTAGATG
SEQ3001	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3002	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3003	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3004	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3005	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3007	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3008	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3009	CTTCAGGACTTTTCCAAACGATGCCAGGTTGGGGTTCAACAGCTACAGTTCAGGATCAAG
SEQ3010	CCTGAGAAGGTTT-----GGAGCCTGTTTACATCTTCTACTTTTGATTAGATGTCGC
SEQ3001	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC--
SEQ3002	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC--
SEQ3003	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC--
SEQ3004	-----
SEQ3005	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC--
SEQ3007	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTAC--
SEQ3008	TAATTCAGCTATTAAAGCTT-----
SEQ3009	TAATTCAGCTATTAAAGCTTATCGTGCTCAAGGTTTATCAGCTTGGGGTTA---
SEQ3010	TTAGTCA-TTTTTGATTTTTTGGCTACGCGAAGTTTATCTGCTTTTGACAAAGA

**Table 30: Comparative Sequences relating to SAG2147
(pr tein of unknown function / lipoprotein, putative)**

>SEQ ID NO 3050: 25_1169NT frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEVPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYKPAQHQTSGQVLSNGNTAGAV
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKAYRAQGLSAWGY

>SEQ ID NO 3051: 25_18RS21 frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKAYRAQGLSAWGY

>SEQ ID NO 3052: 25_2603 frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKAYRAQGLSAWGY

>SEQ ID NO 3053: 25_090 frame: 3

AKKSKMIKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTASEEAAVEQAVV
TENTPATSSQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI GSAAAAQMAAATGVPQST
WEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQ

>SEQ ID NO 3054: 25_A909 frame: 1

KATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTASEEAAVEQAVVTENTPAT
SQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI GSAAAAQMAAATGVPQSTWEHIIAR
ESNGNPNVANASGASGLFQTMPGWGSTATVQDQVNSAIKAYRAQGLS

>SEQ ID NO 3055: 25_CJB110 frame: 3

SLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTAS
EEAAVEQAVVTENTPATSSQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI GSAAAAQMA
AATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVNSAIKAYRA
QGLSAWGY

>SEQ ID NO 3056: 25_COH1 frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYKPAQHQTSGQVLSNGNTAGAV
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKAYRAQGLSAWGY

>SEQ ID NO 3057: 25_H36B frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYRPAQHQTSGQVLSNGNTAGAI
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKA

>SEQ ID NO 3058: 25_M732 frame: 1

KSSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SQQQVTASEEAAVEQAVVTENTPATSSQAQQTAYVTETTYKPAQHQTSGQVLSNGNTAGAV
GSAAAAQMAAATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVN
SAIKAYRAQGLSAWGY

>SEQ ID NO 3059: 25_M781 frame: 4

SLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVTAS
EEAAVEQAVVTENTPATSSQAQQTAYVTETTYKPAQHQTSGQVLSNGNTAGAV GSAAAAQMA
AATGVPQSTWEHIIARESNNGNPNVANASGASGLFQTMPGWGSTATVQDQVNSAIKAYRA
QGLSAWGY

SEQ3050	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEVPKSSSQSTEAN
SEQ3051	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3052	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3053	-----AKKSKMIKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3054	-----KATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3055	-----SLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3056	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3057	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3058	SSQVTTESLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN
SEQ3059	-----SLSKADKVRVAKKSKMTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEAN

பெரிய செய்தி

7

**Table 31: Comparative Sequences relating to SAG2148
(LysM domain protein)**

SEQ ID NO. 3101: SAG2148 FROM THE 1169NT1 GBS TYPE V STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCTGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3102: SAG2148 FROM THE 18RS21 GBS TYPE II STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGTTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3103: SAG2148 FROM THE 2603 V/R GBS TYPE V STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGTTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3104: SAG2148 FROM THE 090 GBS TYPE Ia STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTAAAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGTTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3105: SAG2148 FROM THE A909 GBS TYPE Ia STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3106: SAG2148 FROM THE CJB110 GBS NONTYPEABLE STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTAAAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGTTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3107: SAG2148 FROM THE COH1 GBS TYPE III STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAATAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCAATGAATTCTTCATCAAATTTGAGTTCAAGTGATTGAGCTGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

**Tabl 31: Comparativ Sequences relating to SAG2148
(LysM domain protein)**

**SEQ ID NO. 3108: SAG2148 FROM THE H36b GBS TYPE Ib STRAIN
(REVERSE COMPLEMENT)**

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCATGAATTTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

**SEQ ID NO. 3109: SAG2148 FROM THE JM9130013 GBS TYPE VIII STRAIN
(REVERSE COMPLEMENT)**

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAAGAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGACGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAAGTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCATGAATTTCTTCATCAAATTTGAGTTCAAGTGATTGAGCCGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ ID NO. 3110: SAG2148 FROM THE M732 GBS TYPE III STRAIN

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAATAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCATGAATTTCTTCATCAAATTTGAGTTCAAGTGATTGAGCTGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

**SEQ ID NO. 3111: SAG2148 FROM THE M781 GBS TYPE III STRAIN
(REVERSE COMPLEMENT)**

GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACTACGGTACAATAGTTAGTG
TCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGTGATGTTTTAAATTTGGATAATTCTACAGCTAGTCAA
GCAGAAGCAAAATCTCAACCAACAATTGAAAATTCATGAATTTCTTCATCAAATTTGAGTTCAAGTGATTGAGCTGCA
AAAGAAGAAATAGCTCGTCGTGAATCAAATGGTAGTTTACTGCACAGAATGGACAATATTATGGAAGATATCAACTG
TCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAAGTAGCGGACAATTATGTGGCTTCTCGTTAC
GGATCTTGGTCGGCAGCGCTATCATTTTGAATAGTAACGGCTGGTAT

SEQ3101	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3102	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3103	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3104	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3105	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3106	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3107	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3108	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3109	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3110	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT
SEQ3111	GCATCTTATACCGTGAAATCAGGTGATACCTTATCAGCTATTGCTAAAAATCATAAACT

SEQ3101	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3102	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3103	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3104	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3105	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3106	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3107	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3108	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3109	ACGGTACAAGAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3110	ACGGTACAATAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT
SEQ3111	ACGGTACAATAGTTAGTGTCTCTCAATAGTATCAGTAACGCTGATGTCATCAGTATAGGT

**Table 31: Comparative Sequences relating t SAG2148
(LysM domain protein)**

SEQ3101	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3102	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3103	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3104	GATGTTTTAAAATTGGATAATTCTAAAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3105	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3106	GATGTTTTAAAATTGGATAATTCTAAAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3107	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3108	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3109	GATGTTTTAAAATTGGATAATTCTACAACTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3110	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3111	GATGTTTTAAAATTGGATAATTCTACAGCTAGTCAAGCAGAAGCAAAATCTCAACCAACA
SEQ3101	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3102	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3103	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3104	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3105	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3106	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3107	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3108	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3109	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3110	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3111	ATTGAAAATTCAATGAATTCCTCATCAAATTTGAGTTCAGTGATTGAGCTGCAAAAGAA
SEQ3101	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3102	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3103	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3104	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3105	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3106	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3107	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3108	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3109	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3110	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3111	GAAATAGCTCGTCGTGAATCAAATGGTAGTTTATACTGCACAGAATGGACAATATTATGGA
SEQ3101	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3102	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3103	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3104	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3105	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3106	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3107	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3108	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3109	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3110	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3111	AGATATCAACTGTCTCAATCTTACCTAAATGGCGACTTATCTCCTGAAAATCAAGAAAAA
SEQ3101	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3102	GTAGCGGACAATTATGTGGTTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3103	GTAGCGGACAATTATGTGGTTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3104	GTAGCGGACAATTATGTGGTTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3105	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3106	GTAGCGGACAATTATGTGGTTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3107	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3108	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3109	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3110	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG
SEQ3111	GTAGCGGACAATTATGTGGCTTCTCGTTACGGATCTTGGTCGGCAGCGCTATCATTTTGG

**Table 31: Comparative Sequences relating to SAG2148
(LysM domain protein)**

SEQ3101	AATAGTAACGGCTGGTAT
SEQ3102	AATAGTAACGGCTGGTAT
SEQ3103	AATAGTAACGGCTGGTAT
SEQ3104	AATAGTAACGGCTGGTAT
SEQ3105	AATAGTAACGGCTGGTAT
SEQ3106	AATAGTAACGGCTGGTAT
SEQ3107	AATAGTAACGGCTGGTAT
SEQ3108	AATAGTAACGGCTGGTAT
SEQ3109	AATAGTAACGGCTGGTAT
SEQ3110	AATAGTAACGGCTGGTAT
SEQ3111	AATAGTAACGGCTGGTAT

>SEQ ID NO 3150:15_1169NT frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3151:15_18RS21 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVVSRYGWSAALSFWNSNGWY

>SEQ ID NO 3152:15_2603 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVVSRYGWSAALSFWNSNGWY

>SEQ ID NO 3153:15_090 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSKASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVVSRYGWSAALSFWNSNGWY

>SEQ ID NO 3154:15_A909 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3155:15_CJB110 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSKASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVVSRYGWSAALSFWNSNGWY

>SEQ ID NO 3156:15_COH1 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQ.LVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3157:15_H36B frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3158:15_JM9130013 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVVISIGDVLKLDNSTTSQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3159:15_M732 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQ.LVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

>SEQ ID NO 3160:15_M781 frame: 1
 ASYTVKSGDTLSAIAKNHKTTVQ.LVSLNSISNADVVISIGDVLKLDNSTASQAEAKSQPT
 IENSMNSSSNLSSSDSAAKEEIIARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
 VADNYVASRYGWSAALSFWNSNGWY

**Table 31: Comparative Sequences relating to SAG2148
(LysM domain protein)**

SEQ3150	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3151	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3152	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3153	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3154	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3155	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3156	ASYTVKSGDTLSAIAKNHKTTVQ-LVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3157	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3158	ASYTVKSGDTLSAIAKNHKTTVQELVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3159	ASYTVKSGDTLSAIAKNHKTTVQ-LVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3160	ASYTVKSGDTLSAIAKNHKTTVQ-LVSLNSISNADVISIGDVLKLDNSTASQAEAKSQPT
SEQ3150	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3151	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3152	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3153	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3154	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3155	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3156	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3157	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3158	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3159	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3160	IENSMNSSLSSSDSAAKEE IARRESNGSYTAQNGQYYGRYQLSQSYLNGDLSPENQEK
SEQ3150	VADNYVASRYGWSAALSFWNSNGWY
SEQ3151	VADNYVVSRYGWSAALSFWNSNGWY
SEQ3152	VADNYVVSRYGWSAALSFWNSNGWY
SEQ3153	VADNYVVSRYGWSAALSFWNSNGWY
SEQ3154	VADNYVASRYGWSAALSFWNSNGWY
SEQ3155	VADNYVVSRYGWSAALSFWNSNGWY
SEQ3156	VADNYVASRYGWSAALSFWNSNGWY
SEQ3157	VADNYVASRYGWSAALSFWNSNGWY
SEQ3158	VADNYVASRYGWSAALSFWNSNGWY
SEQ3159	VADNYVASRYGWSAALSFWNSNGWY
SEQ3160	VADNYVASRYGWSAALSFWNSNGWY

Table 32: Conversion of ORF Ref Nos. with SAG Ref N s.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotation
ORF00003	SAG0017	447	PcsB protein
ORF00004	SAG0018	322	ribose-phosphate pyrophosphokinase
ORF00005	SAG0019	391	aminotransferase, class I
ORF00006	SAG0020	253	recombination protein O
ORF00008	SAG0021	283	protease, putative
ORF00009	SAG0022	330	fatty acid/phospholipid synthesis protein PlsX
ORF00010	SAG0023	79	acyl carrier protein
ORF00011	SAG0024	234	phosphoribosylaminoimidazole-succinocarboxamide synthase
ORF00012	SAG0025	1241	phosphoribosylformylglycinamide synthase, putative
ORF00013	SAG0026	484	amidophosphoribosyltransferase
ORF00014	SAG0027	340	phosphoribosylformylglycinamide cyclo-ligase
ORF00015	SAG0028	182	phosphoribosylglycinamide formyltransferase
ORF00016	SAG0029	250	acetyltransferase, GNAT family
ORF00017	SAG0030	515	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase
ORF00018	SAG0031	283	peptidase, M23/M37 family
ORF00020	SAG0032	434	group B streptococcal surface immunogenic protein
ORF00021	SAG0033	232	N-acetylmannosamine-6-P epimerase, putative
ORF00022	SAG0034	438	sugar ABC transporter, sugar-binding protein
ORF00023	SAG0035	295	sugar ABC transporter, permease protein
ORF00024	SAG0036	276	sugar ABC transporter, permease protein
ORF00025	SAG0037	147	conserved hypothetical protein
ORF00026	SAG0038	220	conserved hypothetical protein
ORF00027	SAG0039	305	N-acetylneuraminate lyase, putative
ORF00028	SAG0040	293	ROK family protein
ORF00029	SAG0041	325	acetyl xylan esterase, putative
ORF00030	SAG0042	267	phosphosugar-binding transcriptional regulator, RpiR family, putative
ORF00031	SAG0043	421	phosphoribosylamine-glycine ligase
ORF00032	SAG0044	162	phosphoribosylaminoimidazole carboxylase, catalytic subunit
ORF00033	SAG0045	363	phosphoribosylaminoimidazole carboxylase, ATPase subunit
ORF00035	SAG0046	463	hypothetical protein
ORF00036	SAG0047	432	adenylosuccinate lyase
ORF00037	SAG0048	303	transcriptional regulator, Cro/Ci family
ORF00038	SAG0049	332	Holliday junction DNA helicase RuvB
ORF00039	SAG0050	145	phosphotyrosine protein phosphatase, low molecular weight
ORF00040	SAG0051	126	MORN motif family protein
ORF00041	SAG0052	592	membrane protein, putative
ORF00042	SAG0053	880	aldehyde-alcohol dehydrogenase
ORF00043	SAG0054	338	alcohol dehydrogenase, propanol-preferring
ORF00044	SAG0055	496	threonine synthase
ORF00045	SAG0056	412	MATE efflux family protein
ORF00046	SAG0057	102	ribosomal protein S10
ORF00047	SAG0058	208	ribosomal protein L3
ORF00048	SAG0059	207	ribosomal protein L4
ORF00049	SAG0060	98	ribosomal protein L23
ORF00050	SAG0061	277	ribosomal protein L2
ORF00052	SAG0062	92	ribosomal protein S19
ORF00054	SAG0063	114	ribosomal protein L22
ORF00055	SAG0064	217	ribosomal protein S3

Tabl 32: Conversion of ORF Ref N s. with SAG R f N s.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotation
ORF00056	SAG0065	137	ribosomal protein L16
ORF00058	SAG0066	68	ribosomal protein L29
ORF00059	SAG0067	86	ribosomal protein S17
ORF00060	SAG0068	122	ribosomal protein L14
ORF00061	SAG0069	101	ribosomal protein L24
ORF00063	SAG0070	180	ribosomal protein L5
ORF00064	SAG0071	61	ribosomal protein S14, putative
ORF00065	SAG0072	132	ribosomal protein S8
ORF00066	SAG0073	178	ribosomal protein L6
ORF00068	SAG0074	118	ribosomal protein L18
ORF00069	SAG0075	164	ribosomal protein S5
ORF00070	SAG0076	59	ribosomal protein L30
ORF00071	SAG0077	146	ribosomal protein L15
ORF00072	SAG0078	434	preprotein translocase, SecY subunit
ORF00073	SAG0079	212	adenylate kinase
ORF00074	SAG0080	72	translation initiation factor IF-1
ORF00075	SAG0081	38	ribosomal protein L36
ORF00077	SAG0082	121	ribosomal protein S13
ORF00078	SAG0083	118	ribosomal protein S11
ORF00080	SAG0084	312	DNA-directed RNA polymerase, alpha subunit
ORF00081	SAG0085	128	ribosomal protein L17
ORF00087	SAG0086	97	hypothetical protein
ORF00088	SAG0087	59	hypothetical protein
ORF00089	SAG0088	56	hypothetical protein
ORF00090	SAG0089	183	conserved hypothetical protein
ORF00091	SAG0090	139	conserved hypothetical protein
ORF00093	SAG0091	144	transcriptional regulator ComX1, putative
ORF00094	SAG0092	230	phosphoglycerate mutase family protein
ORF00095	SAG0093	250	D-alanyl-D-alanine carboxypeptidase family protein
ORF00096	SAG0094	191	N-acetylmuramoyl-L-alanine amidase, family 4 protein
ORF00097	SAG0095	344	heat-inducible transcription repressor HrcA
ORF00098	SAG0096	190	heat shock protein GrpE
ORF00099	SAG0097	609	dnaK protein
ORF00100	SAG0098	379	dnaJ protein
ORF00101	SAG0099	415	transcriptional regulator, GntR family
ORF00102	SAG0100	258	tRNA pseudouridine synthase A
ORF00103	SAG0101	252	phosphomethylpyrimidine kinase, putative
ORF00104	SAG0102	154	conserved hypothetical protein
ORF00105	SAG0103	189	conserved hypothetical protein
ORF00106	SAG0104	280	conserved hypothetical protein
ORF00107	SAG0105	427	trigger factor
ORF00108	SAG0106	191	DNA-directed RNA polymerase, delta subunit, putative
ORF00109	SAG0107	534	CTP synthase
ORF00110	SAG0108	308	conserved hypothetical protein
ORF00111	SAG0109	148	deoxyuridine 5'-triphosphate nucleotidohydrolase
ORF00112	SAG0110	454	DNA repair protein RadA
ORF00113	SAG0111	165	carbonic anhydrase-related protein
ORF00115	SAG0112	439	pyridine nucleotide-disulphide oxidoreductase family protein
ORF00116	SAG0113	484	glutamyl-tRNA synthetase
ORF00117	SAG0114	322	ribose ABC transporter, periplasmic D-ribose-binding protein

Tabl 32: Conv rsion of ORF Ref Nos. with SAG R f Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF00118	SAG0115	310	ribose ABC transporter, permease protein
ORF00119	SAG0116	492	ribos ABC transporter, ATP-binding protein
ORF00120	SAG0117	132	ribose ABC transporter protein RbsD
ORF00121	SAG0118	303	ribokinase
ORF00122	SAG0119	328	ribose operon repressor RbsR
ORF00123	SAG0120	32	hypothetical protein
ORF00124	SAG0121	362	permease, putative
ORF00125	SAG0122	228	ABC transporter, ATP-binding protein
ORF00126	SAG0123	223	DNA-binding response regulator
ORF00128	SAG0124	356	sensor histidine kinase
ORF00129	SAG0125	396	argininosuccinate synthase
ORF00130	SAG0126	462	argininosuccinate lyase
ORF00131	SAG0127	293	fructose-bisphosphate aldolase
ORF00132	SAG0128	305	L-2-hydroxyisocaproate dehydrogenase
ORF00133	SAG0129	62	ribosomal protein L28
ORF00134	SAG0130	121	conserved hypothetical protein
ORF00135	SAG0131	543	DAK2 domain protein
ORF00136	SAG0132	294	SPFH domain/Band 7 family protein
ORF00137	SAG0133	38	conserved hypothetical protein
ORF00138	SAG0134	96	hypothetical protein
ORF00141	SAG0135	246	amino acid ABC transporter, ATP-binding protein
ORF00142	SAG0136	516	amino acid ABC transporter, amino acid-binding protein/permease protein
ORF00143	SAG0137	627	conserved hypothetical protein
ORF00145	SAG0138	279	undecaprenol kinase, putative
ORF00146	SAG0139	251	negative regulator of competence MecA, putative
ORF00148	SAG0140	386	glycosyl transferase, group 4 family protein
ORF00149	SAG0141	256	ABC transporter, ATP-binding protein
ORF00150	SAG0142	420	conserved hypothetical protein
ORF00151	SAG0143	410	selenocysteine lyase
ORF00152	SAG0144	147	NifU family protein
ORF00153	SAG0145	472	conserved hypothetical protein
ORF00154	SAG0146	395	penicillin-binding protein 4, putative
ORF00155	SAG0147	411	D-alanyl-D-alanine carboxypeptidase
ORF00156	SAG0148	551	oligopeptide ABC transporter, substrate binding protein, putative
ORF00157	SAG0149	304	oligopeptide ABC transporter, permease protein
ORF00158	SAG0150	343	oligopeptide ABC transporter, permease protein
ORF00160	SAG0151	348	oligopeptide ABC transporter, ATP-binding protein
ORF00161	SAG0152	310	oligopeptide ABC transporter, ATP-binding protein
ORF00166	SAG0153	283	4-diphosphocytidyl-2C-methyl-D-erythritol kinase
ORF00167	SAG0154	147	adc operon repressor AdcR
ORF00168	SAG0155	236	zinc ABC transporter, ATP-binding protein
ORF00169	SAG0156	270	zinc ABC transporter, permease protein
ORF00172	SAG0158	419	tyrosyl-tRNA synthetase
ORF00173	SAG0159	765	penicillin-binding protein 1B, putative
ORF00174	SAG0160	1191	DNA-directed RNA polymerase, beta subunit
ORF00176	SAG0161	1216	DNA-directed RNA polymerase beta' subunit
ORF00178	SAG0162	121	conserved hypothetical protein
ORF00179	SAG0163	323	competence protein CglA
ORF00180	SAG0164	282	competence protein CglB
ORF00181	SAG0165	151	conserved hypothetical protein
ORF00182	SAG0166	123	conserved domain protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref N .	aa	Annotation
ORF00183	SAG0167	324	conserved hypothetical protein
ORF00184	SAG0168	397	acetate kinase
ORF00186	SAG0169	68	transcriptional regulator, Cro/C1 family
ORF00187	SAG0170	45	hypothetical protein
ORF00188	SAG0171	151	hypothetical protein
ORF00189	SAG0172	221	protease, putative
ORF00190	SAG0173	256	pyrroline-5-carboxylate reductase
ORF00191	SAG0174	355	glutamyl-aminopeptidase
ORF00192	SAG0175	79	hypothetical protein
ORF00193	SAG0176	94	conserved hypothetical protein
ORF00194	SAG0177	107	thioredoxin family protein
ORF00195	SAG0178	208	tRNA binding domain protein
ORF00196	SAG0179	238	conserved hypothetical protein
ORF00198	SAG0180	131	single-strand binding protein
ORF00199	SAG0181	214	hydrolase, haloacid dehalogenase-like family
ORF00200	SAG0182	581	sensor histidine kinase, putative
ORF00201	SAG0183	246	response regulator
ORF00203	SAG0184	151	conserved hypothetical protein
ORF00204	SAG0185	242	membrane protein, putative
ORF00205	SAG0186	36	hypothetical protein
ORF00206	SAG0187	542	oligopeptide ABC transporter, oligopeptide-binding protein
ORF00207	SAG0188	325	oligopeptide ABC transporter, permease protein
ORF00208	SAG0189	273	oligopeptide ABC transporter, permease protein
ORF00209	SAG0190	267	peptide ABC transporter, ATP-binding protein
ORF00210	SAG0191	208	peptide ABC transporter, ATP-binding protein
ORF00211	SAG0192	676	PTS system, IIBC components
ORF00212	SAG0193	541	alpha amylase family protein
ORF00214	SAG0194	639	transcriptional antiterminator, BglG family
ORF00216	SAG0195	377	IS1548, transposase
ORF00217	SAG0196	66	conserved domain protein
ORF00218	SAG0197	94	PTS system, IIB component, putative
ORF00219	SAG0198	451	PTS system, IIC component, putative
ORF00220	SAG0199	285	transketolase, N-terminal subunit
ORF00221	SAG0200	309	transketolase, C-terminal subunit
ORF00223	SAG0201	419	oxidoreductase, putative
ORF00224	SAG0202	89	ribosomal protein S15
ORF00225	SAG0203	709	polyribonucleotide nucleotidyltransferase
ORF00226	SAG0204	250	conserved hypothetical protein
ORF00227	SAG0205	194	serine O-acetyltransferase
ORF00228	SAG0206	60	hypothetical protein
ORF00229	SAG0207	447	cysteinyl-tRNA synthetase
ORF00230	SAG0208	128	conserved hypothetical protein
ORF00231	SAG0209	251	RNA methyltransferase, TrmH family, group 3
ORF00232	SAG0210	172	conserved hypothetical protein
ORF00233	SAG0211	286	DegV family protein
ORF00234	SAG0212	32	hypothetical protein
ORF00235	SAG0213	39	hypothetical protein
ORF00236	SAG0214	148	ribosomal protein L13
ORF00237	SAG0215	130	ribosomal protein S9
ORF00238	SAG0216	33	hypothetical protein
ORF00239	SAG0217	384	site-specific recombinase, phage integrase family
ORF00240	SAG0218	158	transcriptional regulator, Cro/C1 family
ORF00241	SAG0219	101	hypothetical protein

Tabl 32: Conv rsion f ORF Ref Nos. with SAG R f Nos.

ORF Ref No.	SAGxxxx R f No.	aa	Annotati n
ORF00242	SAG0220	92	conserved hypothetical protein
ORF00243	SAG0221	76	hypothetical protein
ORF00244	SAG0222	108	conserved domain protein
ORF00245	SAG0223	209	conserved hypothetical protein, fusion
ORF00246	SAG0224	332	replication initiation protein, putative
ORF00247	SAG0225	144	hypothetical protein
ORF00248	SAG0226	418	recombination protein
ORF00249	SAG0227	156	hypothetical protein
ORF00250	SAG0228	111	conserved hypothetical protein
ORF00251	SAG0229	95	conserved hypothetical protein
ORF00252	SAG0230	96	conserved hypothetical protein
ORF00253	SAG0231	135	hypothetical protein
ORF00254	SAG0232	186	hypothetical protein
ORF00255	SAG0233	226	hypothetical protein
ORF00256	SAG0234	128	hypothetical protein
ORF00257	SAG0235	93	hypothetical protein
ORF00258	SAG0236	32	hypothetical protein
ORF00259	SAG0237	34	hypothetical protein
ORF00260	SAG0238	41	hypothetical protein
ORF00261	SAG0239	286	transcriptional regulator MutR family
ORF00262	SAG0240	393	transporter, putative
ORF00263	SAG0241	213	amino acid ABC transporter, permease protein
ORF00264	SAG0242	308	amino acid ABC transporter, amino acid-binding protein
ORF00265	SAG0243	211	amino acid ABC transporter, permease protein
ORF00266	SAG0244	381	amino acid ABC transporter, ATP-binding protein
ORF00272	SAG0245	152	hypothetical protein
ORF00273	SAG0246	268	hypothetical protein
ORF00274	SAG0247	116	hypothetical protein
ORF00275	SAG0248	90	hypothetical protein
ORF00276	SAG0249	116	hypothetical protein
ORF00278	SAG0250	193	hypothetical protein
ORF00279	SAG0251	72	transcriptional regulator, Cro/Ci family
ORF00280	SAG0252	186	acetyltransferase, GNAT family
ORF00281	SAG0253	192	acetyltransferase, GNAT family
ORF00282	SAG0254	226	acetyltransferase, GNAT family
ORF00283	SAG0255	315	conserved hypothetical protein
ORF00284	SAG0256	163	RNA polymerase sigma factor, ECF subfamily
ORF00285	SAG0257	53	hypothetical protein
ORF00287	SAG0258	202	transcriptional regulator, TetR family
ORF00288	SAG0259	365	ABC transporter efflux protein, DrrB family, putative
ORF00289	SAG0260	238	ABC transporter, ATP-binding protein
ORF00290	SAG0261	129	IS1381, transposase OrfB
ORF00291	SAG0262	127	IS1381, transposase OrfA
ORF00292	SAG0263	171	hypothetical protein
ORF00293	SAG0264	103	conserved hypothetical protein
ORF00294	SAG0265	235	conserved hypothetical protein
ORF00295	SAG0266	382	N-acetylglucosamine-6-phosphate deacetylase
ORF00296	SAG0267	180	conserved hypothetical protein
ORF00297	SAG0268	304	glycyl-tRNA synthetase, alpha subunit
ORF00298	SAG0269	213	acyl carrier protein phosphodiesterase, putative
ORF00299	SAG0270	679	glycyl-tRNA synthetase, beta subunit
ORF00300	SAG0271	85	conserved hypothetical protein
ORF00301	SAG0272	87	membrane protein, putative

Tabl 32: Conversion of ORF Ref Nos. with SAG R f Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotati n
ORF00302	SAG0273	502	glycerol kinase
ORF00303	SAG0274	609	alpha-glycerophosphate xidase
ORF00304	SAG0275	232	glycerol uptake facilitator protein
ORF00305	SAG0276	445	NADH oxidase, putative
ORF00306	SAG0277	476	conserved hypothetical protein
ORF00307	SAG0278	661	transketolase
ORF00308	SAG0279	101	conserved hypothetical protein
ORF00309	SAG0280	244	ABC transporter, ATP-binding protein
ORF00310	SAG0281	534	membrane protein, putative
ORF00313	SAG0282	461	PTS system, IIBC components
ORF00314	SAG0283	267	glutamate 5-kinase
ORF00315	SAG0284	417	gamma-glutamyl phosphate reductase
ORF00316	SAG0285	298	conserved hypothetical protein TIGR00006
ORF00317	SAG0286	108	cell division protein FtsL, putative
ORF00318	SAG0287	752	penicillin-binding protein 2X
ORF00319	SAG0288	336	phospho-N-acetylmuramoyl-pentapeptide-transferase
ORF00320	SAG0289	447	ATP-dependent RNA helicase, DEAD/DEAH box family
ORF00321	SAG0290	270	ABC transporter, substrate-binding protein
ORF00322	SAG0291	267	amino acid ABC transporter, permease protein
ORF00323	SAG0292	247	amino acid ABC transporter, ATP-binding protein
ORF00324	SAG0293	74	conserved hypothetical protein
ORF00325	SAG0294	304	thioredoxin reductase
ORF00326	SAG0295	486	conserved hypothetical protein
ORF00327	SAG0296	273	NAD synthetase
ORF00328	SAG0297	444	aminopeptidase C
ORF00329	SAG0298	750	penicillin-binding protein 1A
ORF00330	SAG0299	199	recombination protein U
ORF00331	SAG0300	172	conserved hypothetical protein
ORF00332	SAG0301	40	hypothetical protein
ORF00333	SAG0302	110	conserved hypothetical protein
ORF00335	SAG0303	384	conserved hypothetical protein
ORF00336	SAG0304	487	conserved hypothetical protein
ORF00337	SAG0305	160	autoinducer-2 production protein LuxS
ORF00338	SAG0306	535	KH domain protein
ORF00340	SAG0307	33	hypothetical protein
ORF00341	SAG0308		ABC transporter, ATP-binding protein, FRAMESHIFT
ORF00343	SAG0309	246	ABC transporter, permease protein, putative
ORF00344	SAG0310	361	conserved hypothetical protein
ORF00345	SAG0311		DNA-binding response regulator POINT MUTATION
ORF00347	SAG0312	234	conserved hypothetical protein
ORF00348	SAG0313	209	guanylate kinase
ORF00349	SAG0314	104	DNA-directed RNA polymerase, omega subunit, putative
ORF00350	SAG0315	796	primosomal protein N'
ORF00351	SAG0316	311	methionyl-tRNA formyltransferase
ORF00352	SAG0317	440	Sun protein
ORF00353	SAG0318	245	serine/threonine phosphatase, putative
ORF00354	SAG0319	651	serine/threonine protein kinase
ORF00355	SAG0320	231	conserved hypothetical protein
ORF00356	SAG0321	339	sensor histidine kinase, putative
ORF00358	SAG0322	213	DNA-binding response regulator

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxx Ref No.	aa	Annotation
ORF00359	SAG0323	466	hydrolase, haloacid dehalogenase family/peptidyl-prolyl cis-trans isomerase, cyclophilin type
ORF00360	SAG0324	124	general stress protein, putative
ORF00361	SAG0325	258	pyruvate formate-lyase-activating enzyme
ORF00362	SAG0326	251	transcriptional regulator, DeoR family
ORF00363	SAG0327	327	transcriptional regulator, putative
ORF00364	SAG0328	107	PTS system, cellobiose-specific IIA component
ORF00366	SAG0329	106	PTS system, cellobiose-specific IIB component
ORF00367	SAG0330	433	PTS system, cellobiose-specific IIC component
ORF00368	SAG0331	818	formate acetyltransferase
ORF00369	SAG0332	222	transaldolase family protein
ORF00371	SAG0333	362	glycerol dehydrogenase
ORF00372	SAG0334	308	cysteine synthase A
ORF00373	SAG0335	214	conserved hypothetical protein TIGR00257
ORF00374	SAG0336	429	helicase, putative
ORF00375	SAG0337	221	competence protein F, putative
ORF00376	SAG0338	184	ribosomal subunit interface protein
ORF00382	SAG0339	450	aspartate kinase family protein
ORF00383	SAG0340	216	hydrolase, haloacid dehalogenase-like family
ORF00384	SAG0341	49	hypothetical protein
ORF00385	SAG0342	263	enoyl-CoA hydratase/isomerase family protein
ORF00386	SAG0343	144	transcriptional regulator, MarR family
ORF00387	SAG0344	323	3-oxoacyl-(acyl-carrier-protein) synthase III
ORF00388	SAG0345	74	acyl carrier protein
ORF00390	SAG0346	319	enoyl-(acyl-carrier-protein) reductase II
ORF00391	SAG0347	308	malonyl CoA-acyl carrier protein transacylase
ORF00392	SAG0348	244	3-oxoacyl-[acyl-carrier protein] reductase
ORF00393	SAG0349	410	3-oxoacyl-(acyl-carrier-protein) synthase II
ORF00394	SAG0350	166	acetyl-CoA carboxylase, biotin carboxyl carrier protein
ORF00395	SAG0351	140	(3R)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase
ORF00396	SAG0352	456	acetyl-CoA carboxylase, biotin carboxylase
ORF00397	SAG0353	291	acetyl-CoA carboxylase, carboxyl transferase, beta subunit
ORF00398	SAG0354	257	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
ORF00399	SAG0355	210	conserved hypothetical protein
ORF00400	SAG0356	425	seryl-tRNA synthetase
ORF00402	SAG0357	330	hypothetical protein
ORF00403	SAG0358	120	conserved hypothetical protein
ORF00404	SAG0359	303	PTS system, mannose-specific IID component
ORF00405	SAG0360	270	PTS system, mannose-specific IIC component
ORF00406	SAG0361	336	PTS system, mannose-specific IIB components
ORF00407	SAG0362	270	hydrolase, haloacid dehalogenase-like family
ORF00408	SAG0363	194	hypothetical protein
ORF00409	SAG0364	203	membrane protein, putative
ORF00410	SAG0365	473	xanthine/uracil permease family protein
ORF00411	SAG0366	169	conserved hypothetical protein TIGR00150, putative
ORF00412	SAG0367	186	acetyltransferase, GNAT family
ORF00413	SAG0368	435	transcriptional regulator, putative
ORF00414	SAG0369	98	conserved hypothetical protein
ORF00415	SAG0370	139	HIT family protein
ORF00416	SAG0371	167	hypothetical protein
ORF00417	SAG0372	85	hypothetical protein

Tabl 32: C nv rsion f ORF R f Nos. with SAG Ref Nos.

ORF Ref N .	SAGxxxx R f No.	aa	Ann tati n
ORF00419	SAG0373	241	ABC transporter, ATP-binding protein
ORF00421	SAG0374	344	ABC transporter, permease protein
ORF00422	SAG0375	266	conserved hypothetical protein
ORF00423	SAG0376	211	conserved hypothetical protein TIGR00091
ORF00424	SAG0377	127	conserved hypothetical protein, POINT MUTATION
ORF00425	SAG0378	379	N utilization substance protein A
ORF00426	SAG0379	98	conserved hypothetical protein
ORF00427	SAG0380	100	ribosomal protein L7A family
ORF00428	SAG0381	927	translation initiation factor IF-2
ORF00429	SAG0382	122	ribosome-binding factor A
ORF00430	SAG0383	334	conserved hypothetical protein
ORF00431	SAG0384	138	transcriptional repressor CopY
ORF00432	SAG0385	744	copper-transporter ATPase CopA
ORF00433	SAG0386	68	copper-transporter protein CopZ
ORF00434	SAG0387	204	conserved hypothetical protein
ORF00435	SAG0388	270	hydrolase, haloacid dehalogenase-like family
ORF00436	SAG0389	880	DNA polymerase I
ORF00437	SAG0390	146	CoA binding domain protein
ORF00438	SAG0391	159	transcriptional regulator, Fur family
ORF00439	SAG0392	521	cell wall surface anchor family protein
ORF00440	SAG0393	228	DNA-binding response regulator
ORF00441	SAG0394	345	sensor histidine kinase
ORF00442	SAG0395	246	conserved hypothetical protein
ORF00443	SAG0396	380	queuine tRNA-ribosyltransferase
ORF00444	SAG0397	102	conserved hypothetical protein
ORF00445	SAG0398	179	bioY family protein
ORF00446	SAG0399	258	AtsA/ElaC family protein
ORF00447	SAG0400	168	cytidine/deoxycytidylate deaminase family protein
ORF00448	SAG0401	44	hypothetical protein
ORF00449	SAG0402	449	glucose-6-phosphate isomerase
ORF00450	SAG0403	175	5-formyltetrahydrofolate cyclo-ligase family protein
ORF00451	SAG0404	225	rhomboid family protein
ORF00452	SAG0405	347	lipoprotein
ORF00453	SAG0406	299	UTP-glucose-1-phosphate uridylyltransferase
ORF00454	SAG0407	338	glycerol-3-phosphate dehydrogenase (NAD(P)+)
ORF00455	SAG0408	109	ribonuclease P protein component
ORF00456	SAG0409	271	SpolIJ family protein
ORF00458	SAG0410	273	R3H domain protein
ORF00463	SAG0411	177	conserved hypothetical protein
ORF00464	SAG0412	258	RecX protein
ORF00465	SAG0413	451	RNA methyltransferase, TrmA family
ORF00466	SAG0414	153	conserved hypothetical protein
ORF00467	SAG0415	142	acetyltransferase, GNAT family
ORF00468	SAG0416	1233	protease, putative
ORF00469	SAG0417	302	glycosyl transferase, group 2 family protein
ORF00470	SAG0418	336	ribonucleoside-diphosphate reductase 2, beta subunit
ORF00471	SAG0419	137	nrdI protein
ORF00472	SAG0420	721	ribonucleoside-diphosphate reductase 2, alpha subunit
ORF00473	SAG0421	1055	conserved hypothetical protein
ORF00474	SAG0422	129	conserved hypothetical protein
ORF00475	SAG0423	132	conserved domain protein

Tabl 32: Conv rsi n of ORF Ref Nos. with SAG Ref Nos.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotati n
ORF00476	SAG0424	94	hypothetical protein
ORF00478	SAG0425	105	carboxymuconolactone decarboxylase family protein
ORF00479	SAG0426	131	conserved hypothetical protein
ORF00480	SAG0427	129	transcriptional regulator, MerR family
ORF00482	SAG0428	345	alcohol dehydrogenase, zinc-containing
ORF00483	SAG0429	284	oxidoreductase, aldo/keto reductase family
ORF00484	SAG0430	287	cation efflux system protein
ORF00485	SAG0431	174	transcriptional regulator, TetR family
ORF00486	SAG0432	397	transcriptional regulator, AraC family
ORF00487	SAG0433	1389	surface protein Rib
ORF00488	SAG0434	61	transposase, IS256 family, truncation
ORF00489	SAG0435	97	DNA-damage-inducible protein J, putative
ORF00490	SAG0436	62	hypothetical protein
ORF00491	SAG0437	123	hypothetical protein
ORF00493	SAG0438	145	bacteriophage L54a, integrase, truncation
ORF00495	SAG0439		conserved hypothetical protein, FRAMESHIFT
ORF00496	SAG0440	84	conserved hypothetical protein
ORF00497	SAG0441	103	conserved domain protein
ORF00499	SAG0442	189	acetyltransferase, GNAT family
ORF00500	SAG0443	194	acetyltransferase, GNAT family
ORF00501	SAG0444	188	conserved hypothetical protein
ORF00502	SAG0445	883	valyl-tRNA synthetase
ORF00503	SAG0446	319	oxidoreductase, Gfo/ldh/MocA family
ORF00504	SAG0447	287	magnesium transporter, CorA family
ORF00506	SAG0448	391	transposase, IS256 family
ORF00507	SAG0449	354	conserved hypothetical protein
ORF00508	SAG0450	330	aspartate-ammonia ligase
ORF00510	SAG0451	149	bacteriocin transport accessory protein, putative
ORF00511	SAG0452	179	type II DNA modification methyltransferase, putative
ORF00512	SAG0453	98	hypothetical protein
ORF00513	SAG0454	161	phosphopantetheine adenylyltransferase
ORF00515	SAG0455	357	conserved hypothetical protein
ORF00518	SAG0456		conserved hypothetical protein, degenerate
ORF00519	SAG0457	192	conserved hypothetical protein
ORF00520	SAG0458	368	conserved hypothetical protein TIGR00048
ORF00521	SAG0459	171	VanZF domain protein
ORF00522	SAG0460	581	ABC transporter, ATP-binding/permease protein
ORF00523	SAG0461	579	ABC transporter, ATP-binding/permease protein
ORF00524	SAG0462	188	anthranilate synthase component II
ORF00525	SAG0463	179	bioY family protein
ORF00526	SAG0464	330	biotin synthetase
ORF00527	SAG0465	164	hypothetical protein
ORF00528	SAG0466	371	thiolase
ORF00531	SAG0467	409	AMP-binding enzyme domain protein
ORF00532	SAG0468	210	endonuclease III
ORF00533	SAG0469	131	type IV prepilin peptidase-related protein
ORF00534	SAG0470	69	conserved hypothetical protein
ORF00535	SAG0471	322	glucokinase
ORF00536	SAG0472	126	rhodanese domain protein
ORF00537	SAG0473	613	elongation factor Tu family protein
ORF00538	SAG0474	81	conserved hypothetical protein
ORF00540	SAG0475	451	UDP-N-acetylmuramoylalanine-D-glutamate ligase

Tabl 32: C nversi n of ORF R f Nos. with SAG Ref N s.

ORF Ref No.	SAGxxxx Ref No.	aa	Ann tation
ORF00541	SAG0476	358	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
ORF00542	SAG0477	378	cell division protein DivIB, putative
ORF00544	SAG0478	429	cell division protein FtsA
ORF00545	SAG0479	426	cell division protein FtsZ
ORF00546	SAG0480	224	ylmE protein, putative
ORF00547	SAG0481	201	ylmF protein
ORF00548	SAG0482	84	YGGT family protein
ORF00549	SAG0483	262	ylmH protein
ORF00550	SAG0484	256	cell division protein DivIVA, putative
ORF00552	SAG0485	930	isoleucyl-tRNA synthetase
ORF00553	SAG0486	100	conserved hypothetical protein
ORF00554	SAG0487	151	MutT/nudix family protein
ORF00555	SAG0488	753	ATP-dependent Clp protease, ATP-binding subunit
ORF00556	SAG0489	34	hypothetical protein
ORF00557	SAG0490	76	conserved hypothetical protein
ORF00558	SAG0491	230	amino acid ABC transporter, permease protein
ORF00559	SAG0492	244	amino acid ABC transporter, ATP-binding protein
ORF00560	SAG0493	564	phosphoglucomutase/phosphomannomutase family protein
ORF00562	SAG0494	284	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
ORF00563	SAG0495	278	conserved hypothetical protein
ORF00564	SAG0496	446	exodeoxyribonuclease VII, large subunit
ORF00565	SAG0497	71	exodeoxyribonuclease VII, small subunit
ORF00566	SAG0498	290	geranyltranstransferase, putative
ORF00567	SAG0499	275	hemolysin A
ORF00568	SAG0500	157	arginine repressor ArgR, putative
ORF00570	SAG0501	552	DNA repair protein RecN
ORF00571	SAG0502	278	DegV family protein
ORF00572	SAG0503	279	Lipase/Acylhydrolase, putative
ORF00573	SAG0504	200	conserved hypothetical protein
ORF00574	SAG0505	91	DNA-binding protein HU
ORF00575	SAG0506	65	hypothetical protein
ORF00576	SAG0507	310	dihydroorotate dehydrogenase A
ORF00577	SAG0508	411	beta-lactam resistance factor
ORF00578	SAG0509	403	beta-lactam resistance factor
ORF00579	SAG0510	406	murM protein, putative
ORF00580	SAG0511	270	hydrolase, haloacid dehalogenase-like family
ORF00581	SAG0512	438	HD domain protein
ORF00582	SAG0513	128	conserved hypothetical protein
ORF00583	SAG0514	894	cation-transporting ATPase, E1-E2 family
ORF00584	SAG0515	286	conserved hypothetical protein
ORF00585	SAG0516	643	fructose-1,6-bisphosphatase, putative
ORF00586	SAG0517	374	iron-sulfur cluster-binding protein, putative
ORF00587	SAG0518		peptide chain release factor 2, FRAMESHIFT
ORF00588	SAG0519	230	cell division ABC transporter, ATP-binding protein FtsE
ORF00589	SAG0520	309	cell division ABC transporter, permease protein FtsX
ORF00590	SAG0521	236	carboxymethylenebutenolldase-related protein
ORF00591	SAG0522	232	metallo-beta-lactamase superfamily protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF R f No.	SAGxxxx R f No.	aa	Annotation
ORF00592	SAG0523	254	oxidoreductase, short chain dehydrogenase/reductase family
ORF00593	SAG0524	835	DNA polymerase III, epsilon subunit/ATP-dependent helicase DinG
ORF00595	SAG0525	397	aspartate aminotransferase
ORF00596	SAG0526	448	asparaginyl-tRNA synthetase
ORF00597	SAG0527	185	conserved hypothetical protein
ORF00598	SAG0528	327	inosine-uridine preferring nucleoside hydrolase
ORF00599	SAG0529	38	hypothetical protein
ORF00600	SAG0530	137	OsmC/Ohr family protein
ORF00601	SAG0531	296	conserved hypothetical protein
ORF00602	SAG0532	324	conserved hypothetical protein
ORF00603	SAG0533	303	Uncharacterized BCR, COG1481
ORF00604	SAG0534	465	dipeptidase
ORF00605	SAG0535	506	zinc ABC transporter, zinc-binding adhesion liprotein
ORF00606	SAG0536	86	ribosomal protein L31
ORF00607	SAG0537	311	DHH family protein
ORF00608	SAG0538	340	adenosine deaminase, putative
ORF00609	SAG0539	147	flavodoxin
ORF00610	SAG0540	91	chorismate mutase, putative
ORF00611	SAG0541	398	voltage-gated chloride channel family protein
ORF00612	SAG0542	127	IS1381, transposase OrfA
ORF00613	SAG0543	129	IS1381, transposase OrfB
ORF00614	SAG0544	115	ribosomal protein L19
ORF00615	SAG0545	359	site-specific recombinase, phage integrase family
ORF00617	SAG0546	67	conserved domain protein
ORF00618	SAG0547	185	hypothetical protein
ORF00619	SAG0548	265	repressor protein, putative
ORF00620	SAG0549	47	hypothetical protein
ORF00621	SAG0550	74	conserved hypothetical protein
ORF00622	SAG0551	52	conserved hypothetical protein
ORF00623	SAG0552	62	hypothetical protein
ORF00624	SAG0553	268	hypothetical protein
ORF00626	SAG0554	63	transcriptional regulator, Cro/C1 family
ORF00627	SAG0555	249	antirepressor, putative
ORF00628	SAG0556	47	hypothetical protein
ORF00630	SAG0557	76	hypothetical protein
ORF00632	SAG0558	74	hypothetical protein
ORF00633	SAG0559	286	conserved hypothetical protein
ORF00634	SAG0560	77	conserved hypothetical protein
ORF00635	SAG0561	46	hypothetical protein
ORF00636	SAG0562	84	hypothetical protein
ORF00637	SAG0563	53	hypothetical protein
ORF00638	SAG0564	160	conserved hypothetical protein
ORF00639	SAG0565	224	conserved domain protein
ORF00640	SAG0566	138	single-strand binding protein
ORF00641	SAG0567	439	reverse transcriptase/maturase family protein
ORF00642	SAG0568	67	conserved hypothetical protein
ORF00643	SAG0569	158	conserved hypothetical protein
ORF00644	SAG0570	115	hypothetical protein
ORF00645	SAG0571	43	hypothetical protein
ORF00646	SAG0572	138	conserved hypothetical protein
ORF00647	SAG0573	54	hypothetical protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref N.	SAGxxxx Ref N.	aa	Annotation
ORF00648	SAG0574	89	conserved hypothetical protein
ORF00649	SAG0575	110	hypothetical protein
ORF00650	SAG0576	43	hypothetical protein
ORF00652	SAG0577	177	conserved hypothetical protein
ORF00653	SAG0578	88	conserved hypothetical protein
ORF00654	SAG0581	118	conserved hypothetical protein
ORF00655	SAG0582	422	conserved hypothetical protein
ORF00656	SAG0583	406	conserved hypothetical protein
ORF00657	SAG0584	62	conserved hypothetical protein, truncation
ORF00658	SAG0585	471	conserved hypothetical protein
ORF00659	SAG0586	154	conserved hypothetical protein
ORF00660	SAG0587	300	structural protein, putative
ORF00661	SAG0588	71	conserved hypothetical protein
ORF00662	SAG0589	143	conserved hypothetical protein
ORF00663	SAG0590	112	conserved hypothetical protein
ORF00664	SAG0591	78	conserved hypothetical protein
ORF00665	SAG0592	111	conserved hypothetical protein
ORF00666	SAG0593	185	structural protein
ORF00667	SAG0594	81	conserved hypothetical protein
ORF00668	SAG0595	123	conserved hypothetical protein
ORF00669	SAG0596	670	PblA, internal deletion
ORF00670	SAG0597	506	minor structural protein, putative
ORF00671	SAG0598	1374	minor structural protein, putative
ORF00672	SAG0599	668	minor structural protein, putative
ORF00673	SAG0600	109	hypothetical protein
ORF00674	SAG0601	70	hypothetical protein
ORF00675	SAG0602	100	conserved hypothetical protein
ORF00676	SAG0603	111	holin, putative
ORF00677	SAG0604	239	lysine, putative
ORF00678	SAG0605	323	conserved hypothetical protein
ORF00679	SAG0606	66	conserved hypothetical protein
ORF00681	SAG0607	56	conserved hypothetical protein
ORF00682	SAG0608	59	hypothetical protein
ORF00683	SAG0609	193	site-specific recombinase, phage integrase family
ORF00685	SAG0610	134	conserved hypothetical protein
ORF00687	SAG0611		transposase, degenerate FRAMESHIFT
ORF00689	SAG0612	53	conserved hypothetical protein, FRAMESHIFT
ORF00690	SAG0613	425	transmembrane protein Vexp1
ORF00691	SAG0614	218	ABC transporter, ATP-binding protein Vexp2
ORF00692	SAG0615	458	transmembrane protein Vexp3
ORF00693	SAG0616	217	DNA-binding response regulator VncR
ORF00694	SAG0617	439	sensor histidine kinase VncS
ORF00695	SAG0618	195	transposase OrfB, IS3 family, truncation
ORF00697	SAG0619	66	conserved hypothetical protein
ORF00698	SAG0620	62	hypothetical protein
ORF00699	SAG0621	401	rod shape-determining protein RodA, putative
ORF00700	SAG0622	186	hydrolase, haloacid dehalogenase-like family
ORF00701	SAG0623	650	DNA gyrase, B subunit
ORF00702	SAG0624	574	septation ring formation regulator EzrA, putative
ORF00703	SAG0625	213	phosphoserine phosphatase SerB
ORF00704	SAG0626	161	MutT/nudix family protein
ORF00705	SAG0627	151	conserved hypothetical protein
ORF00706	SAG0628	435	enolase
ORF00707	SAG0629	354	conserved domain protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF00708	SAG0630	427	3-phosphoshikimate 1-carboxyvinyltransferase
ORF00709	SAG0631	170	shikimate kinase
ORF00710	SAG0632	457	psr protein
ORF00711	SAG0633	451	RNA methyltransferase, TrmA family
ORF00712	SAG0634	70	hypothetical protein
ORF00713	SAG0635	245	acid phosphatase precursor, class B
ORF00714	SAG0636	172	conserved hypothetical protein
ORF00717	SAG0637		transcriptional regulator, TetR family, putative, FRAMESHIFT
ORF00718	SAG0638	109	cell wall surface anchor family protein
ORF00720	SAG0639	273	transposase OrfB, IS3 family
ORF00721	SAG0640	91	transposase OrfA, IS3 family
ORF00722	SAG0641		Tn5252, Orf 10 protein, degenerate POINT MUTATION
ORF00723	SAG0642	59	hypothetical protein
ORF00725	SAG0643		chaperonin, 33 kDa DEGENERATE
ORF00726	SAG0644	402	transcriptional regulator, AraC family
ORF00727	SAG0645	554	cell wall surface anchor family protein, putative
ORF00728	SAG0646	307	cell wall surface anchor family protein
ORF00729	SAG0647	305	sortase family protein
ORF00731	SAG0648	260	sortase family protein
ORF00732	SAG0649	890	cell wall surface anchor family protein, putative
ORF00734	SAG0650	189	sortase family protein, FRAMESHIFT
ORF00735	SAG0651	201	hypothetical protein
ORF00737	SAG0653	76	conserved hypothetical protein, DEGENERATE
ORF00738	SAG0654	34	hypothetical protein
ORF00740	SAG0656	36	hypothetical protein
ORF00741	SAG0657	89	hypothetical protein
ORF00742	SAG0658	383	lipoprotein, putative
ORF00743	SAG0659	330	ABC transporter, ATP-binding protein
ORF00744	SAG0660	272	membrane protein
ORF00745	SAG0661	261	conserved hypothetical protein
ORF00747	SAG0663	282	cylD protein
ORF00748	SAG0664	240	cylG protein
ORF00749	SAG0665	101	acyl carrier protein AcpC
ORF00750	SAG0666	158	cylZ protein FRAMESHIFT
ORF00751	SAG0667	309	cylA protein
ORF00752	SAG0668	292	cylB protein
ORF00753	SAG0669	667	cylE protein
ORF00754	SAG0670	317	cylF protein
ORF00755	SAG0671	731	cylI protein
ORF00756	SAG0672	403	cylJ protein
ORF00757	SAG0673	191	cylK protein
ORF00758	SAG0674	113	hypothetical protein
ORF00759	SAG0675	171	surface protein antigen-related protein
ORF00760	SAG0676	885	serine protease, subtilase family, putative
ORF00761	SAG0677	1062	hypothetical protein
ORF00762	SAG0678		endopeptidase O DEGENERATE
ORF00766	SAG0679	286	hydrolase, alpha/beta fold family, putative
ORF00767	SAG0680	339	hypothetical protein
ORF00768	SAG0681	353	conserved domain protein
ORF00769	SAG0682	409	permease, putative
ORF00770	SAG0683		transmembrane protein Vexp3, putative FRAMESHIFT
ORF00774	SAG0684	223	ABC transporter, ATP-binding protein

Table 32: Conversion of ORF Ref N s. with SAG Ref N s.

ORF Ref N .	SAGxxxx R f No.	aa	Annotation
ORF00775	SAG0685	472	conserved hypothetical protein
ORF00776	SAG0686	261	DNA-entry nuclease, putative
ORF00777	SAG0687	212	DedA family protein, putative
ORF00778	SAG0688	218	ABC transporter, ATP-binding protein
ORF00779	SAG0689	257	membrane protein, putative
ORF00780	SAG0690	272	conserved hypothetical protein
ORF00781	SAG0691	294	transcriptional regulator, LysR family
ORF00783	SAG0692	193	regulatory protein, putative
ORF00785	SAG0693	377	IS1548, transposase
ORF00786	SAG0694	173	regulatory protein, putative, truncation
ORF00787	SAG0695	330	D-lactate dehydrogenase
ORF00788	SAG0696	516	sodium:galactoside symporter family protein, putative
ORF00789	SAG0697	341	2-keto-3-deoxygluconate kinase
ORF00790	SAG0698	599	beta-glucuronidase
ORF00791	SAG0699	223	transcriptional regulator, GntR family
ORF00792	SAG0700	205	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase
ORF00793	SAG0701	466	Glucuronate isomerase
ORF00794	SAG0702	348	mannonate dehydratase
ORF00795	SAG0703	279	D-mannonate oxidoreductase
ORF00796	SAG0704	270	hydrolase, haloacid dehalogenase-like family
ORF00797	SAG0705	596	glycosyl hydrolase, family 3
ORF00798	SAG0706	361	proline dipeptidase
ORF00799	SAG0707	334	transcriptional regulator, RegM family
ORF00800	SAG0708	488	alpha amylase family protein
ORF00801	SAG0709	332	glycosyl transferase, group 1 family protein
ORF00802	SAG0710	444	glycosyl transferase, group 1 family protein
ORF00803	SAG0711	647	threonyl-tRNA synthetase
ORF00804	SAG0712	234	DNA-binding response regulator
ORF00805	SAG0713	339	conserved hypothetical protein
ORF00806	SAG0714	188	conserved hypothetical protein
ORF00807	SAG0715	216	amino acid ABC transporter, permease protein
ORF00808	SAG0716	231	amino acid ABC transporter, permease protein
ORF00809	SAG0717	266	amino acid ABC transporter, amino acid-binding protein
ORF00810	SAG0718	251	amino acid ABC transporter, ATP-binding protein
ORF00811	SAG0719	236	DNA-binding response regulator
ORF00812	SAG0720	449	sensory box histidine kinase
ORF00813	SAG0721	269	metallo-beta-lactamase family protein
ORF00814	SAG0722	122	conserved hypothetical protein
ORF00815	SAG0723	236	ribonuclease III
ORF00816	SAG0724	1179	SMC family protein
ORF00817	SAG0725	265	hydrolase, haloacid dehalogenase-like family
ORF00818	SAG0726	274	hydrolase, haloacid dehalogenase-like family
ORF00819	SAG0727	536	signal recognition particle-docking protein FtsY
ORF00820	SAG0728	270	ABC transporter, substrate-binding protein
ORF00821	SAG0729	300	ABC transporter, permease protein, putative
ORF00822	SAG0730	42	ABC transporter, ATP-binding protein
ORF00823	SAG0731	347	bacterial luciferase family protein
ORF00824	SAG0732	720	transcriptional accessory protein Tex, putative
ORF00825	SAG0733	142	conserved hypothetical protein
ORF00826	SAG0734	87	phage shock protein C, putative
ORF00827	SAG0735	44	hypothetical protein
ORF00828	SAG0736	311	HPr(Ser) kinase/phosphatase

Tabl 32: Conv rsion f ORF Ref Nos. with SAG R f Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF00830	SAG0737	257	prolipoprotein diacylglycerol transferase
ORF00832	SAG0738	132	conserved hypothetical protein
ORF00833	SAG0739	143	conserved hypothetical protein
ORF00834	SAG0740	91	conserved hypothetical protein
ORF00835	SAG0741	303	peptidase, U32 family, putative
ORF00836	SAG0742	428	peptidase, U32 family
ORF00837	SAG0743	70	conserved hypothetical protein
ORF00838	SAG0744	265	membrane protein, putative
ORF00839	SAG0745	446	Mn2+/Fe2+ transporter, NRAMF family
ORF00840	SAG0746	369	riboflavin biosynthesis protein RibD
ORF00841	SAG0747	208	riboflavin synthase, alpha subunit
ORF00842	SAG0748	397	riboflavin biosynthesis protein RibA
ORF00843	SAG0749	156	riboflavin synthase, beta subunit
ORF00844	SAG0750	496	lysyl-tRNA synthetase
ORF00845	SAG0751	300	hydrolase, haloacid dehalogenase-like family
ORF00846	SAG0752	213	phosphoglycerate mutase family protein
ORF00847	SAG0753	157	ebsC family protein, putative
ORF00848	SAG0754	205	conserved domain protein
ORF00850	SAG0765	282	peptidase, U32 family
ORF00852	SAG0766	174	conserved hypothetical protein
ORF00853	SAG0767	129	lipoprotein, putative
ORF00855	SAG0768	599	oligoendopeptidase F, putative
ORF00856	SAG0769	931	phosphoenolpyruvate carboxylase
ORF00857	SAG0760	377	IS1548, transposase
ORF00859	SAG0761	422	cell division protein, FtsW/RodA/SpoVE family
ORF00861	SAG0762	398	translation elongation factor Tu
ORF00863	SAG0763	252	triosephosphate isomerase
ORF00865	SAG0764	230	phosphoglycerate mutase
ORF00866	SAG0765	681	penicillin-binding protein 2b
ORF00867	SAG0766	198	recombination protein RecR
ORF00868	SAG0767	348	D-alanine-D-alanine ligase
ORF00869	SAG0768	455	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase
ORF00870	SAG0769	406	oxalate:formate antiporter
ORF00871	SAG0770	228	conserved hypothetical protein
ORF00872	SAG0771	512	cell wall surface anchor family protein
ORF00873	SAG0772	514	peptide chain release factor 3
ORF00874	SAG0773	126	conserved hypothetical protein
ORF00876	SAG0774	244	ABC transporter, ATP-binding protein
ORF00878	SAG0775	220	ABC transporter, permease protein
ORF00879	SAG0776	276	lipoprotein, putative
ORF00880	SAG0777	528	ATP-dependent RNA helicase, DEAD/DEAH box family
ORF00882	SAG0778	88	conserved hypothetical protein
ORF00883	SAG0779	254	conserved hypothetical protein
ORF00884	SAG0780	246	acyltransferase family protein
ORF00885	SAG0781	217	competence protein CeaA
ORF00887	SAG0782	745	DNA internalization-related competence protein ComEC/Rec2
ORF00888	SAG0783	269	hydrolase, haloacid dehalogenase-like family
ORF00889	SAG0784	314	sugar-binding transcriptional regulator, LacI family
ORF00890	SAG0785	330	conserved hypothetical protein
ORF00891	SAG0786	242	conserved domain protein
ORF00892	SAG0787	345	DNA polymerase III, delta subunit, putative

Tabl 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF00893	SAG0788	202	superoxide dismutase, Fe-Mn
ORF00894	SAG0789	283	transcriptional antiterminator LicT
ORF00895	SAG0790	622	PTS system, beta-glucosides-specific IIBC components
ORF00896	SAG0791	475	6-phospho-beta-glucosidase
ORF00898	SAG0792	364	conserved hypothetical protein
ORF00899	SAG0793	380	conserved hypothetical protein TIGR00045
ORF00900	SAG0794	418	permease, GntP family
ORF00902	SAG0795	354	conserved hypothetical protein
ORF00903	SAG0796	147	transcriptional regulator, MarR family
ORF00904	SAG0797	342	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
ORF00905	SAG0798	226	membrane protein, putative
ORF00906	SAG0799	233	glucosamine-6-phosphate isomerase
ORF00907	SAG0800	318	Glutathione S-transferases domain protein
ORF00908	SAG0801	239	ribosomal small subunit pseudouridine synthase
ORF00909	SAG0802	38	hypothetical protein
ORF00910	SAG0803	383	major facilitator family protein
ORF00911	SAG0804	315	competence protein ColA
ORF00912	SAG0805	601	oligoendopeptidase B
ORF00913	SAG0806	208	hydrolase, haloacid dehalogenase-like family
ORF00914	SAG0807	235	O-methyltransferase family protein
ORF00916	SAG0808	309	protease maturation protein, putative
ORF00918	SAG0809	161	conserved hypothetical protein
ORF00919	SAG0810	872	alanine-tRNA synthetase
ORF00921	SAG0811	238	membrane protein, putative
ORF00922	SAG0812	272	glycosyl transferase, family 8
ORF00923	SAG0813	81	hypothetical protein
ORF00924	SAG0814	95	conserved domain protein
ORF00925	SAG0815	71	transcriptional regulator, Cro/C1 family
ORF00926	SAG0816	253	conserved hypothetical protein
ORF00927	SAG0817	187	conserved hypothetical protein
ORF00928	SAG0818	319	ribonucleoside-diphosphate reductase 2, beta subunit
ORF00929	SAG0819	719	ribonucleoside-diphosphate reductase 2, alpha subunit
ORF00930	SAG0820	74	ribonucleoside-diphosphate reductase 2, NrdH-redoxin
ORF00931	SAG0821	87	phosphocarrier protein HPr
ORF00932	SAG0822	577	phosphoenolpyruvate-protein phosphotransferase
ORF00933	SAG0823	475	glyceraldehyde-3-phosphate dehydrogenase, NADP-dependent
ORF00934	SAG0824	417	polysaccharide deacetylase family protein
ORF00935	SAG0825	360	ATP-dependent RNA helicase, DEAD/DEAH box family
ORF00936	SAG0826	209	uridine kinase
ORF00937	SAG0827	165	conserved hypothetical protein
ORF00938	SAG0828	554	DNA polymerase III, gamma and tau subunits
ORF00939	SAG0829	64	conserved hypothetical protein
ORF00940	SAG0830	311	biotin-acetyl-CoA-carboxylase ligase
ORF00941	SAG0831	398	S-adenosylmethionine synthetase
ORF00942	SAG0832	753	hypothetical protein
ORF00943	SAG0833	181	hypothetical protein
ORF00944	SAG0834	42	hypothetical protein
ORF00945	SAG0835	188	conserved hypothetical protein

Tabl 32: Conv rsion of ORF R f Nos. with SAG R f N s.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotation
ORF00946	SAG0836	184	conserved hypothetical protein
ORF00948	SAG0837	428	ABC transporter, ATP-binding protein
ORF00950	SAG0838	233	hypothetical protein
ORF00951	SAG0839	226	transcriptional regulator, TenA family
ORF00952	SAG0840	265	phosphomethylpyrimidine kinase
ORF00953	SAG0841	256	hydroxyethylthiazole kinase
ORF00954	SAG0842	223	thiamine-phosphate pyrophosphorylase
ORF00955	SAG0843	419	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
ORF00956	SAG0844	184	acetyltransferase, GNAT family
ORF00957	SAG0845	427	CBS domain protein
ORF00958	SAG0846	286	methionine aminopeptidase, type I
ORF00959	SAG0847	306	ribonuclease BN, putative
ORF00961	SAG0848	151	GtrA family protein
ORF00962	SAG0849	169	conserved hypothetical protein
ORF00963	SAG0850	652	DNA ligase, NAD-dependent
ORF00964	SAG0851	339	BmrU protein, putative
ORF00966	SAG0852	766	pullulanase, putative
ORF00967	SAG0853	622	1,4-alpha-glucan branching enzyme
ORF00968	SAG0854	379	glucose-1-phosphate adenylyltransferase
ORF00969	SAG0855		glycogen biosynthesis protein GlgD FRAMESHIFT
ORF00971	SAG0856	476	glycogen synthase
ORF00972	SAG0857	66	ATP synthase F0, C subunit
ORF00973	SAG0858	238	ATP synthase F0, A subunit
ORF00974	SAG0859	165	ATP synthase F0, B subunit
ORF00975	SAG0860	178	ATP synthase F1, delta subunit
ORF00976	SAG0861	501	ATP synthase F1, alpha subunit
ORF00977	SAG0862	293	ATP synthase F1, gamma subunit
ORF00978	SAG0863	468	ATP synthase F1, beta subunit
ORF00979	SAG0864	137	ATP synthase F1, epsilon subunit
ORF00980	SAG0865	76	conserved hypothetical protein
ORF00981	SAG0866	423	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
ORF00982	SAG0867	63	conserved hypothetical protein
ORF00983	SAG0868	285	DNA-entry nuclease
ORF00984	SAG0869	346	phenylalanyl-tRNA synthetase, alpha subunit
ORF00985	SAG0870	173	acetyltransferase, GNAT family
ORF00986	SAG0871	801	phenylalanyl-tRNA synthetase, beta subunit
ORF00987	SAG0872	300	conserved hypothetical protein
ORF00988	SAG0873	1077	exonuclease RxB
ORF00989	SAG0874	1207	exonuclease RxA
ORF00990	SAG0875	305	magnesium transporter, CorA family, putative
ORF00991	SAG0876	458	tRNA modification GTPase TrmE
ORF00992	SAG0877	636	ABC transporter, ATP-binding protein
ORF00993	SAG0878	322	acetoin dehydrogenase, thymine PPi dependent, E1 component, alpha subunit
ORF00994	SAG0879	332	acetoin dehydrogenase, thymine PPi dependent, E1 component, beta subunit
ORF00995	SAG0880	462	acetoin dehydrogenase, thymine PPi dependent, E2 component, dihydrolipoamide acetyltransferase
ORF00996	SAG0881	585	acetoin dehydrogenase, thymine PPi dependent, E3 component, dihydrolipoamide dehydrogenase
ORF00997	SAG0882	329	lipoate-protein ligase A
ORF00998	SAG0883	261	cobryic acid synthase, putative

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF00999	SAG0884	447	mur ligase family protein
ORF01000	SAG0885	283	conserved hypothetical protein TIGR00159
ORF01001	SAG0886	319	Gram-positive signal peptide, YSIRK family domain protein
ORF01002	SAG0887	450	phosphoglucosyltransferase/phosphomannosyltransferase family protein
ORF01003	SAG0888	123	conserved hypothetical protein
ORF01004	SAG0889	126	conserved hypothetical protein
ORF01005	SAG0890	376	oxygen-independent coproporphyrinogen III oxidase, putative
ORF01006	SAG0891	245	conserved hypothetical protein
ORF01007	SAG0892	256	hydrolase, haloacid dehalogenase-like family
ORF01008	SAG0893	218	conserved hypothetical protein
ORF01009	SAG0894	1370	conserved hypothetical protein
ORF01010	SAG0895	289	lipoyl-binding domain protein
ORF01011	SAG0896	108	oxidoreductase, putative
ORF01012	SAG0897	221	conserved hypothetical protein
ORF01013	SAG0898	83	hypothetical protein
ORF01014	SAG0899	57	hypothetical protein
ORF01015	SAG0900	56	hypothetical protein
ORF01016	SAG0901	127	hypothetical protein
ORF01018	SAG0902	45	hypothetical protein
ORF01019	SAG0903	44	hypothetical protein
ORF01021	SAG0904	56	hypothetical protein
ORF01022	SAG0905	138	nucleoside diphosphate kinase
ORF01023	SAG0906	610	GTP-binding protein LepA
ORF01024	SAG0907	877	streptococcal histidine triad family protein
ORF01025	SAG0908	203	HD domain protein
ORF01026	SAG0909	154	acetyltransferase, GNAT family
ORF01027	SAG0910	144	PilB-related protein
ORF01030	SAG0911	930	cation-transporting ATPase, E1-E2 family
ORF01031	SAG0912	367	nucleoside diphosphate kinase domain protein
ORF01032	SAG0913	212	chloramphenicol acetyltransferase
ORF01033	SAG0914	203	conserved hypothetical protein
ORF01034	SAG0915	405	Tn916, transposase
ORF01035	SAG0916	67	Tn916, excisionase
ORF01037	SAG0918	76	Tn916, hypothetical protein
ORF01038	SAG0919	157	Tn916, hypothetical protein
ORF01039	SAG0921	117	Tn916, transcriptional regulator, putative
ORF01040	SAG0923	639	Tn916, tetracycline resistance protein
ORF01041	SAG0925	310	Tn916, hypothetical protein
ORF01042	SAG0926	333	Tn916, NLP/P60 family protein
ORF01044	SAG0927	725	Tn916, hypothetical protein FRAMESHIFT
ORF01047	SAG0928		Tn916, hypothetical protein FRAMESHIFT
ORF01048	SAG0929	168	Tn916, hypothetical protein
ORF01049	SAG0930	165	Tn916, hypothetical protein
ORF01050	SAG0931	73	Tn916, hypothetical protein
ORF01051	SAG0932	401	Tn916, transcriptional regulator, putative
ORF01052	SAG0933	461	Tn916, FtsK/SpoIIIE family protein
ORF01053	SAG0934	128	Tn916, hypothetical protein
ORF01054	SAG0935	104	Tn916, hypothetical protein
ORF01056	SAG0937		ABC transporter, ATP-binding protein, FRAMESHIFT
ORF01057	SAG0938	122	transcriptional regulator, GntR family
ORF01058	SAG0939	1034	DNA polymerase III, alpha subunit

Tabl 32: Conversion of ORF Ref N s. with SAG Ref N s.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotation
ORF01059	SAG0940	340	6-phosph fructokinase
ORF01060	SAG0941	500	pyruvate kinase
ORF01061	SAG0942	185	signal peptidase I, putative
ORF01062	SAG0943	47	hypothetical protein
ORF01063	SAG0944	604	glucosamine--fructose-6-phosphate aminotransferase (isomerizing)
ORF01064	SAG0945	377	IS1548, transposase
ORF01066	SAG0946	109	phnA protein
ORF01068	SAG0947	213	amino acid ABC transporter, permease protein
ORF01069	SAG0948	209	amino acid ABC transporter, ATP-binding protein
ORF01070	SAG0949	276	amino acid ABC transporter, amino acid-binding protein
ORF01072	SAG0950	82	ribosomal protein S20
ORF01073	SAG0951	306	pantothenate kinase
ORF01074	SAG0952	196	conserved hypothetical protein
ORF01075	SAG0953	129	cytidine deaminase
ORF01076	SAG0954	349	lipoprotein
ORF01077	SAG0955	511	sugar ABC transporter, ATP-binding protein
ORF01078	SAG0956	353	sugar ABC transporter, permease protein, putative
ORF01079	SAG0957	318	sugar ABC transporter, permease protein, putative
ORF01080	SAG0958	456	NADH oxidase
ORF01081	SAG0959	329	L-lactate dehydrogenase
ORF01082	SAG0960	819	DNA gyrase, A subunit
ORF01083	SAG0961	247	sortase SrtA
ORF01084	SAG0962	137	glyoxylase family protein
ORF01085	SAG0963	320	conserved hypothetical protein
ORF01086	SAG0964	375	Na ⁺ /H ⁺ exchanger family protein
ORF01087	SAG0965	127	IS1381, transposase OrfA
ORF01088	SAG0966	129	IS1381, transposase OrfB
ORF01089	SAG0967	520	GMP synthase
ORF01090	SAG0968	232	transcriptional regulator, GntR family
ORF01091	SAG0969	444	gid protein
ORF01092	SAG0970	247	acetyltransferase, GNAT family
ORF01093	SAG0971	282	lipoprotein, putative
ORF01095	SAG0972		conserved hypothetical protein, FRAMESHIFT
ORF01096	SAG0973	320	nisin-resistance protein, putative
ORF01097	SAG0974	250	ABC transporter, ATP-binding protein
ORF01098	SAG0975	651	ABC transporter, permease protein, putative
ORF01099	SAG0976	222	DNA-binding response regulator
ORF01100	SAG0977	312	sensor histidine kinase
ORF01101	SAG0978	356	site-specific recombinase, phage integrase family
ORF01102	SAG0979	553	ABC transporter, substrate binding protein, putative
ORF01103	SAG0980	257	conserved hypothetical protein
ORF01104	SAG0981	228	SatD
ORF01106	SAG0982	521	signal recognition particle protein
ORF01108	SAG0983	110	conserved hypothetical protein
ORF01109	SAG0984	437	sensor histidine kinase CiaH
ORF01110	SAG0985	226	DNA-binding response regulator CiaR
ORF01111	SAG0986	849	aminopeptidase N
ORF01112	SAG0987	217	phosphate transport system regulatory protein PhoU

Table 32: Comparison of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01113	SAG0988	252	phosphate ABC transporter, ATP-binding protein PstB, putative
ORF01114	SAG0989	267	phosphate ABC transporter, ATP-binding protein PstB, putative
ORF01115	SAG0990	295	phosphate ABC transporter, permease protein PstA, putative
ORF01116	SAG0991	305	phosphate ABC transporter, permease protein
ORF01117	SAG0992	286	phosphate ABC transporter, phosphate-binding protein
ORF01118	SAG0993	436	NOL1/NOP2/sun family protein
ORF01119	SAG0994	254	inositol monophosphatase family protein
ORF01120	SAG0995	93	conserved hypothetical protein
ORF01121	SAG0996	137	conserved hypothetical protein
ORF01122	SAG0997	310	macrolide-efflux protein mreA/riboflavin biosynthesis protein RibF
ORF01123	SAG0998	294	tRNA pseudouridine synthase B
ORF01124	SAG0999	143	acetyltransferase, GNAT family
ORF01125	SAG1000	423	conserved hypothetical protein
ORF01126	SAG1001	196	conserved hypothetical protein
ORF01127	SAG1002	292	protease, putative
ORF01128	SAG1003	876	permease, putative
ORF01129	SAG1004	233	ABC transporter, ATP-binding protein
ORF01131	SAG1005	706	DNA topoisomerase I
ORF01132	SAG1006	280	DprA/SMF protein, putative DNA processing factor
ORF01133	SAG1007	342	Iron-compound ABC transporter, iron-compound-binding protein
ORF01134	SAG1008	253	Iron compound ABC transporter, ATP-binding protein
ORF01135	SAG1009	324	Iron compound ABC transporter, permease protein
ORF01136	SAG1010	320	Iron compound ABC transporter, permease protein
ORF01137	SAG1011	182	acetyltransferase, CysE/LacA/LpxA/NodL family
ORF01138	SAG1012	253	ribonuclease HII
ORF01139	SAG1013	283	GTP-binding protein
ORF01140	SAG1014	190	conserved hypothetical protein
ORF01142	SAG1015	494	carbon starvation protein CstA, putative
ORF01143	SAG1016	244	response regulator
ORF01144	SAG1017	579	sensor histidine kinase, putative
ORF01145	SAG1018	40	hypothetical protein
ORF01146	SAG1019	39	conserved hypothetical protein, FRAMESHIFT
ORF01148	SAG1020	227	hypothetical protein
ORF01149	SAG1021	107	hypothetical protein
ORF01150	SAG1022	177	hypothetical protein
ORF01151	SAG1023	48	hypothetical protein
ORF01152	SAG1024	183	hypothetical protein
ORF01153	SAG1025	149	hypothetical protein
ORF01156	SAG1026		immunogenic secreted protein, DEGENERATE
ORF01157	SAG1027	84	conserved hypothetical protein
ORF01158	SAG1028	196	hypothetical protein
ORF01159	SAG1029	101	hypothetical protein
ORF01160	SAG1030	304	conserved hypothetical protein
ORF01161	SAG1031	120	extracellular protein, putative POINT MUTATION
ORF01162	SAG1032	85	conserved hypothetical protein
ORF01164	SAG1033	1309	FtsK/SpoIIIE family protein
ORF01166	SAG1034	55	hypothetical protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01167	SAG1035	424	conserved hypothetical protein
ORF01168	SAG1036	80	conserved hypothetical protein
ORF01169	SAG1037	157	hypothetical protein
ORF01172	SAG1038	1003	phage infection protein, putative
ORF01173	SAG1039	96	conserved hypothetical protein
ORF01174	SAG1040	260	conserved domain protein
ORF01175	SAG1041	107	hypothetical protein
ORF01176	SAG1042	1060	carbamoyl-phosphate synthase, large subunit
ORF01177	SAG1043	358	carbamoyl-phosphate synthase, small subunit
ORF01178	SAG1044	307	aspartate carbamoyltransferase
ORF01179	SAG1045	430	dihydroorotase, multifunctional complex type
ORF01180	SAG1046	209	orotate phosphoribosyltransferase
ORF01181	SAG1047	233	orotidine 5'-phosphate decarboxylase
ORF01182	SAG1048	410	membrane protein, putative
ORF01183	SAG1049	513	ABC transporter, ATP-binding protein
ORF01184	SAG1050	112	ribonucleotide reductase, truncation
ORF01185	SAG1051	358	aspartate-semialdehyde dehydrogenase
ORF01186	SAG1052	47	cell wall surface anchor family protein, putative
ORF01187	SAG1053	30	hypothetical protein
ORF01188	SAG1054	531	cardiolipin synthetase
ORF01189	SAG1055	556	formate-tetrahydrofolate ligase
ORF01190	SAG1056	339	lipoate-protein ligase A
ORF01191	SAG1057	292	conserved hypothetical protein
ORF01192	SAG1058	272	conserved hypothetical protein
ORF01193	SAG1059	110	glycine cleavage system H protein, putative
ORF01194	SAG1060	328	bacterial luciferase family protein
ORF01195	SAG1061	399	oxidoreductase, FMN-binding
ORF01197	SAG1062	282	lipoate-protein ligase A family protein
ORF01198	SAG1063	228	flavoprotein-related protein
ORF01199	SAG1064	180	flavoprotein family protein
ORF01200	SAG1065	190	membrane protein, putative
ORF01201	SAG1066	572	phosphoglucomutase
ORF01202	SAG1067	178	IS861, transposase OrfA
ORF01203	SAG1068	277	IS861, transposase OrfB
ORF01204	SAG1069	65	hypothetical protein
ORF01205	SAG1070	577	ABC transporter, ATP-binding/permease protein
ORF01206	SAG1071	573	ABC transporter, ATP-binding/permease protein
ORF01207	SAG1072	200	conserved hypothetical protein
ORF01208	SAG1073	325	conserved hypothetical protein
ORF01209	SAG1074	418	Serine hydroxymethyltransferase
ORF01210	SAG1075	183	Sua5/YciO/YrdC/YwIC family protein
ORF01211	SAG1076	276	modification methylase, HemK family
ORF01212	SAG1077	359	peptide chain release factor 1
ORF01213	SAG1078	189	thymidine kinases
ORF01214	SAG1079	60	4-oxalocrotonate tautomerase
ORF01215	SAG1080	47	hypothetical protein
ORF01216	SAG1081	312	ApbE family protein
ORF01217	SAG1082	200	conserved hypothetical protein
ORF01218	SAG1083	411	conserved hypothetical protein
ORF01219	SAG1084	262	formate/nitrite transporter family protein
ORF01220	SAG1085	424	xanthine permease
ORF01221	SAG1086	193	xanthine phosphoribosyltransferase
ORF01222	SAG1087	327	guanosine monophosphate reductase
ORF01223	SAG1088	446	drug resistance transporter, EmrB/QacA family, putative

Tabl 32: Conv rsion of ORF R f N s. with SAG R f Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Ann tation
ORF01224	SAG1089	230	conserved hypothetical protein
ORF01225	SAG1090	666	potassium uptak protein, putative
ORF01226	SAG1091	216	oxidoreductase, short chain dehydrogenase/reductase family, FRAMESHIFT
ORF01227	SAG1092	330	phosphate acetyltransferase
ORF01228	SAG1093	294	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF01229	SAG1094	278	conserved hypothetical protein
ORF01230	SAG1095	223	GTP pyrophosphokinase family protein
ORF01231	SAG1096	190	conserved hypothetical protein
ORF01232	SAG1097	324	ribose-phosphate pyrophosphokinase
ORF01233	SAG1098	371	cysteine desulphurase
ORF01234	SAG1099	115	conserved hypothetical protein
ORF01235	SAG1100	210	DNA-binding protein
ORF01236	SAG1101	226	DNA repair protein RadC
ORF01237	SAG1102	377	membrane protein, putative
ORF01238	SAG1103	478	6-phospho-beta-glucosidase
ORF01239	SAG1104	204	platelet activating factor, putative
ORF01240	SAG1105	273	hydrolase, haloacid dehalogenase-like family
ORF01241	SAG1106	309	transcriptional regulator, AraC family, putative
ORF01242	SAG1107	510	voltage-gated chloride channel family protein
ORF01243	SAG1108	357	spermidine/putrescine ABC transporter, spermidine/putrescine-binding protein
ORF01244	SAG1109	258	spermidine/putrescine ABC transporter, permease protein
ORF01245	SAG1110	264	spermidine/putrescine ABC transporter, permease protein
ORF01246	SAG1111	384	spermidine/putrescine ABC transporter, ATP-binding protein
ORF01247	SAG1112	300	UDP-N-acetylenolpyruvoylglucosamine reductase
ORF01248	SAG1113	162	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase
ORF01249	SAG1114	120	dihydroneopterin aldolase
ORF01250	SAG1115	267	dihydropteroate synthase
ORF01251	SAG1116	187	GTP cyclohydrolase I
ORF01252	SAG1117	420	folylpolyglutamate synthase
ORF01253	SAG1118	295	rarD protein
ORF01254	SAG1119	288	homoserine kinase
ORF01255	SAG1120	427	homoserine dehydrogenase
ORF01256	SAG1121	295	polysaccharide deacetylase family protein
ORF01257	SAG1122	515	transporter, BCCT family protein
ORF01258	SAG1123	34	hypothetical protein
ORF01259	SAG1124	458	aldehyde dehydrogenase family protein
ORF01260	SAG1125	335	membrane protein
ORF01261	SAG1126	228	conserved hypothetical protein
ORF01262	SAG1127	113	conserved hypothetical protein, FRAMESHIFT
ORF01263		187	hypothetical protein
ORF01264	SAG1128	65	transcriptional regulator, Cro/C1 family
ORF01265	SAG1129	36	hypothetical protein
ORF01266	SAG1130	49	hypothetical protein
ORF01268	SAG1131	164	thiol peroxidase
ORF01269	SAG1132	219	conserved hypothetical protein
ORF01272	SAG1133	254	cons rved hypothetical protein

Tabl 32: C nversi n of ORF Ref N s. with SAG Ref Nos.

ORF Ref No.	SAGxxxx R f No.	aa	Annotation
ORF01273	SAG1134	213	transcriptional regulator, GntR family/potassium uptake protein, TrkA family
ORF01274	SAG1135	183	gls24 protein, putative
ORF01275	SAG1136		conserved hypothetical protein FRAMESHIFT
ORF01276	SAG1137	180	gls24 protein, putative
ORF01277	SAG1138	64	conserved hypothetical protein
ORF01279	SAG1139	193	conserved hypothetical protein
ORF01280	SAG1140	82	conserved hypothetical protein
ORF01281	SAG1141	112	conserved hypothetical protein
ORF01282	SAG1142	759	ATP-dependent DNA helicase PcrA
ORF01283	SAG1143	100	conserved hypothetical protein, FRAMESHIFT
ORF01284	SAG1144	441	uracil permease
ORF01285	SAG1145	448	sodium:alanine symporter family protein
ORF01286	SAG1146	411	cation efflux family protein
ORF01287	SAG1147	130	conserved hypothetical protein
ORF01288	SAG1148	231	membrane protein, putative
ORF01289	SAG1149	207	conserved hypothetical protein
ORF01290	SAG1150	400	ribosomal protein S1
ORF01291	SAG1151	76	conserved hypothetical protein
ORF01292	SAG1152	340	branched-chain amino acid aminotransferase
ORF01294	SAG1153	819	DNA topoisomerase IV, A subunit
ORF01295	SAG1154	653	DNA topoisomerase IV, B subunit
ORF01296	SAG1155	207	conserved hypothetical protein TIGR00023
ORF01297	SAG1156	217	uracil-DNA glycosylase
ORF01298	SAG1157	161	conserved hypothetical protein
ORF01299	SAG1158	413	CMP-N-acetylneuraminic acid synthetase NeuA
ORF01300	SAG1159	209	neuD protein
ORF01301	SAG1160	384	UDP-N-acetylglucosamine-2-epimerase NeuC
ORF01302	SAG1161	341	N-acetyl neuramic acid synthetase NeuB
ORF01303	SAG1162	466	cpsL protein
ORF01304	SAG1163	318	cpsVK protein
ORF01305	SAG1164	321	cpsVJ protein
ORF01306	SAG1165	327	cpsVO protein
ORF01307	SAG1166	295	cpsVN protein
ORF01308	SAG1167	241	cpsVM protein
ORF01309	SAG1168	364	cpsVH protein
ORF01310	SAG1169	163	CpsVG
ORF01311	SAG1170	149	CpsF
ORF01312	SAG1171	462	CpsE
ORF01313	SAG1172	229	CpsD protein
ORF01314	SAG1173	230	cpsC protein
ORF01315	SAG1174	243	capsular polysaccharide biosynthesis protein CpsB
ORF01316	SAG1175	485	capsular polysaccharide biosynthesis protein CpsA
ORF01317	SAG1176	290	capsular polysaccharide synthesis operon transcriptional regulator CpsY
ORF01318	SAG1177	255	cpsIaS protein
ORF01319	SAG1178	236	purine nucleoside phosphorylase
ORF01320	SAG1179	418	voltage-gated chloride channel family protein, putative
ORF01321	SAG1180	269	purine nucleoside phosphorylase
ORF01322	SAG1181	135	arsenate reductase
ORF01323	SAG1182	403	phosphopentomutase
ORF01324	SAG1183	223	ribose 5-phosphate isomerase

Tabl 32: Conversion of ORF Ref Nos. with SAG R f Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01326	SAG1184	236	conserved hypothetical protein
ORF01327	SAG1185	262	tributylin esterase
ORF01328	SAG1186	553	metallo-beta-lactamase superfamily protein
ORF01329	SAG1187	253	ABC transporter, ATP-binding protein
ORF01330	SAG1188	287	ABC transporter, permease protein
ORF01331	SAG1189	334	conserved hypothetical protein
ORF01332	SAG1190	551	adherence and virulence protein A
ORF01333	SAG1191	239	alpha-acetolactate decarboxylase
ORF01334	SAG1192	560	acetolactate synthase, catabolic
ORF01335	SAG1193	408	TPR domain protein
ORF01336	SAG1194	396	membrane protein
ORF01337	SAG1195	153	MutT/nudix family protein
ORF01338	SAG1196	160	mutator MutT protein
ORF01339	SAG1197	1072	hyaluronidase
ORF01340	SAG1198	348	dTDP-glucose 4,6-dehydratase
ORF01341	SAG1199	197	dTDP-4-dehydrorhamnose 3,5-epimerase
ORF01342	SAG1200	289	glucose-1-phosphate thymidyltransferase
ORF01343	SAG1201	367	iminodiacetate oxidase, putative
ORF01344	SAG1202	262	conserved hypothetical protein TIGR00486
ORF01345	SAG1203	227	conserved hypothetical protein
ORF01346	SAG1204	226	DNA replication protein Dnad, putative
ORF01347	SAG1205	172	adenine phosphoribosyltransferase
ORF01348	SAG1206	854	conserved domain protein
ORF01349	SAG1207	32	hypothetical protein
ORF01350	SAG1208	732	single-stranded-DNA-specific exonuclease RecJ
ORF01351	SAG1209	253	oxidoreductase, short chain dehydrogenase/reductase family
ORF01352	SAG1210	309	metallo-beta-lactamase superfamily protein
ORF01353	SAG1211	215	conserved hypothetical protein
ORF01354	SAG1212	412	GTP-binding protein HflX
ORF01355	SAG1213	296	tRNA delta(2)-isopentenylpyrophosphate transferase
ORF01356	SAG1214	58	hypothetical protein
ORF01357	SAG1215	305	exfoliative toxin A, putative
ORF01358	SAG1216	1252	pullulanase, putative
ORF01361	SAG1217		conserved hypothetical protein, FRAMESHIFT
ORF01362	SAG1218	194	conserved hypothetical protein
ORF01363	SAG1219	468	peptidase, M20/M25/M40 family
ORF01364	SAG1220	200	nitroreductase family protein
ORF01365	SAG1221		glycerophosphoryl diester phosphodiesterase, putative, POINT MUTATION
ORF01367	SAG1222	593	excinuclease ABC, C subunit
ORF01368	SAG1223	255	conserved hypothetical protein
ORF01369	SAG1224	446	MATE efflux family protein
ORF01370	SAG1225	136	conserved hypothetical protein
ORF01371	SAG1226	165	conserved hypothetical protein
ORF01372	SAG1227	198	conserved hypothetical protein
ORF01373	SAG1228	96	ISSdy1, transposase OrfA
ORF01374	SAG1229	259	ISSdy1, transposase OrfB
ORF01375	SAG1230	96	conserved hypothetical protein
ORF01377	SAG1231		transposase OrfB, IS3 family, degenerate FRAMESHIFT
ORF01379	SAG1232	77	transposase OrfB, IS3 family, truncation
ORF01380	SAG1233	822	streptococcal histidine triad family protein
ORF01381	SAG1234	306	laminin-binding surface protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01382	SAG1235	425	GBS11, group II intron, maturase
ORF01383	SAG1236		c5a peptidase precursor FRAMESHIFT
ORF01384	SAG1237	444	hypothetical protein
ORF01385	SAG1238	202	hypothetical protein
ORF01386	SAG1239	76	conserved hypothetical protein
ORF01387	SAG1240	125	conserved hypothetical protein, truncation
ORF01388	SAG1241	78	transposase OrfA, IS3 family
ORF01389	SAG1242	67	transposase OrfB, IS3 family, truncation
ORF01390	SAG1243	96	ISSdy1, transposase OrfA FRAMESHIFT
ORF01391	SAG1244	259	ISSdy1, transposase OrfB
ORF01392	SAG1245	38	hypothetical protein
ORF01393	SAG1246	389	hypothetical protein
ORF01394	SAG1247	399	integrase, phage family
ORF01395	SAG1248	75	conserved hypothetical protein
ORF01396	SAG1249	74	transcriptional regulator, Cro/C1 family
ORF01397	SAG1250	621	Tn5252, relaxase
ORF01398	SAG1251	121	Tn5252, Orf 9 protein
ORF01399	SAG1252	120	Tn5252, Orf 10 protein
ORF01401	SAG1253	435	transposase, ISL3 family
ORF01403	SAG1254	546	mercuric reductase
ORF01404	SAG1255	130	mercuric resistance operon regulatory protein MerR
ORF01406	SAG1256	142	IS861, transposase OrfB, truncation
ORF01407	SAG1257	709	cation-transporting ATPase, E1-E2 family
ORF01408	SAG1258	122	cadmium efflux system accessory protein
ORF01409	SAG1259	99	conserved hypothetical protein
ORF01410	SAG1260	262	hypothetical protein
ORF01411	SAG1261	198	conserved hypothetical protein
ORF01412	SAG1262	695	cation-transporting ATPase, E1-E2 family
ORF01414	SAG1263		conserved domain protein, FRAMESHIFT
ORF01415	SAG1264	148	transcriptional repressor CopY, putative
ORF01416	SAG1265	206	cadmium resistance transporter, putative
ORF01417	SAG1266	152	hypothetical protein
ORF01418	SAG1267	108	hypothetical protein
ORF01419	SAG1268	230	repressor protein, putative
ORF01420	SAG1269	44	hypothetical protein
ORF01421	SAG1270	471	ImpB/MucB/SamB family protein
ORF01423	SAG1271	116	conserved hypothetical protein
ORF01424	SAG1272	102	conserved hypothetical protein
ORF01425	SAG1273	118	conserved hypothetical protein
ORF01426	SAG1274	129	conserved hypothetical protein
ORF01427	SAG1275	75	hypothetical protein
ORF01428	SAG1276	358	conserved hypothetical protein
ORF01430	SAG1277	163	hypothetical protein
ORF01431	SAG1278	96	hypothetical protein
ORF01432	SAG1279	99	conserved domain protein
ORF01433	SAG1280	2274	Helicases conserved C-terminal domain protein
ORF01434	SAG1281	183	hypothetical protein
ORF01435	SAG1282	63	lipoprotein, putative
ORF01436	SAG1283	1631	cell wall surface anchor family protein
ORF01437	SAG1284	196	abortive infection protein AbiGI
ORF01438	SAG1285	281	abortive infection protein AbiGII
ORF01439	SAG1286	933	conserved hypothetical protein
ORF01440	SAG1287	776	conserved hypothetical protein
ORF01441	SAG1288	117	conserved hypothetical protein, DEGENERATE

Tabl 32: Conversion of ORF R f Nos. with SAG Ref N s.

ORF Ref N .	SAGxxxx R f No.	aa	Annotation
ORF01442	SAG1289	284	conserved hypothetical protein
ORF01443	SAG1290	80	hypothetical protein
ORF01444	SAG1291	605	Tn5252, Orf 21 protein, internal deletion
ORF01445	SAG1292	162	hypothetical protein
ORF01446	SAG1293	194	protease, putative
ORF01447	SAG1294	77	conserved hypothetical protein
ORF01449	SAG1295	127	conserved hypothetical protein
ORF01450	SAG1296	142	conserved hypothetical protein
ORF01451	SAG1297	451	type II DNA modification methyltransferase Spn5252IP
ORF01452	SAG1298	31	hypothetical protein
ORF01453	SAG1299	272	conserved hypothetical protein
ORF01454	SAG1300	57	conserved hypothetical protein
ORF01455	SAG1301	121	ribosomal protein L7/L12
ORF01456	SAG1302	166	ribosomal protein L10
ORF01458	SAG1303	702	ATP-dependent Clp protease, ATP-binding subunit
ORF01459	SAG1304	32	hypothetical protein
ORF01460	SAG1305	314	homocysteine S-methyltransferase MmuM, putative
ORF01461	SAG1306	458	amino acid permease
ORF01463	SAG1307	216	hypothetical protein
ORF01464	SAG1308	167	hypothetical protein
ORF01465	SAG1309	30	hypothetical protein
ORF01466	SAG1310	182	transcriptional regulator, TetR family
ORF01467	SAG1311	198	GTP-binding protein
ORF01468	SAG1312	408	ATP-dependent Clp protease, ATP-binding subunit ClpX
ORF01469	SAG1313	56	conserved hypothetical protein
ORF01470	SAG1314	164	dihydrofolate reductase
ORF01471	SAG1315	279	thymidylate synthase
ORF01472	SAG1316	390	HMG-CoA synthase
ORF01473	SAG1317	427	3-hydroxy-3-methylglutaryl-CoA reductase
ORF01474	SAG1318	149	conserved hypothetical protein
ORF01475	SAG1319	187	hemolysin III, putative
ORF01476	SAG1320	304	conserved hypothetical protein TIGR00147
ORF01477	SAG1321	284	glutathione S-transferase family protein
ORF01478	SAG1322	72	conserved domain protein
ORF01479	SAG1323	331	isopentenyl-diphosphate delta-isomerase
ORF01480	SAG1324	330	phosphomevalonate kinase
ORF01481	SAG1325	314	diphosphomevalonate decarboxylase
ORF01482	SAG1326	292	mevalonate kinase, putative
ORF01483	SAG1327	409	sensor histidine kinase
ORF01484	SAG1328	228	DNA-binding response regulator
ORF01485	SAG1329	208	GTP pyrophosphokinase family protein
ORF01486	SAG1330	68	hypothetical protein
ORF01488	SAG1331	979	R5 protein
ORF01489	SAG1332	146	transcriptional regulator, MarR family, putative
ORF01490	SAG1333	690	5'-nucleotidase family protein
ORF01491	SAG1334	136	polypeptide deformylase, putative
ORF01492	SAG1335	449	NADP-specific glutamate dehydrogenase
ORF01494	SAG1336	169	conserved hypothetical protein
ORF01495	SAG1337	589	ABC transporter, ATP-binding/permease protein
ORF01496	SAG1338	579	ABC transporter, ATP-binding/permease protein
ORF01497	SAG1339	157	acetyltransferase, GNAT family

Tabl 32: Conversion of ORF Ref N s. with SAG Ref Nos.

ORF Ref No.	SAGxxxx R f No.	aa	Annotation
ORF01498	SAG1340	622	ABC transporter, ATP-binding protein
ORF01499	SAG1341	402	polyA polymerase family protein
ORF01500	SAG1342	282	DegV family protein
ORF01501	SAG1343	126	conserved hypothetical protein
ORF01502	SAG1344	177	hypothetical protein
ORF01503	SAG1345	164	conserved hypothetical protein
ORF01504	SAG1346	641	PTS system, fructose specific IIBC components
ORF01505	SAG1347	303	1-phosphofructokinase
ORF01506	SAG1348	247	lactose phosphotransferase system repressor
ORF01507	SAG1349	411	beta-lactam resistance factor
ORF01508	SAG1350	544	surface antigen-related protein
ORF01509	SAG1351	307	2-dehydropantoate 2-reductase, putative
ORF01510	SAG1352	356	regulatory protein, putative
ORF01511	SAG1353	330	pyridine nucleotide-disulphide oxidoreductase family protein
ORF01512	SAG1354	251	tRNA (guanine-N1)-methyltransferase
ORF01513	SAG1355	172	16S rRNA processing protein RimM
ORF01515	SAG1356	503	transcriptional regulator, RofA family
ORF01516	SAG1357	80	KH domain protein
ORF01517	SAG1358	90	ribosomal protein S16
ORF01518	SAG1359	415	permease, putative
ORF01519	SAG1360	236	ABC transporter, ATP-binding protein
ORF01520	SAG1361	414	conserved hypothetical protein
ORF01522	SAG1362	532	carbamoyl-phosphate synthase, large subunit, putative
ORF01523	SAG1363	356	carbamoyl-phosphate synthase, small subunit
ORF01524	SAG1364	173	pyrimidine operon regulatory protein
ORF01525	SAG1365	296	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF01526	SAG1366	154	lipoprotein signal peptidase
ORF01527	SAG1367	301	transcriptional regulator, LysR family
ORF01528	SAG1368	94	ribosomal protein L27
ORF01529	SAG1369	112	conserved hypothetical protein
ORF01530	SAG1370	104	ribosomal protein L21
ORF01531	SAG1371	392	conserved hypothetical protein
ORF01532	SAG1372	404	thiamine biosynthesis protein Thil
ORF01533	SAG1373	381	cysteine desulphurase
ORF01535	SAG1374	150	conserved hypothetical protein
ORF01536	SAG1375	449	glutathione reductase
ORF01537	SAG1376	111	conserved hypothetical protein
ORF01538	SAG1377	388	chorismate synthase
ORF01539	SAG1378	355	3-dehydroquinase synthase
ORF01540	SAG1379	225	3-dehydroquinase dehydratase
ORF01541	SAG1380	385	conserved hypothetical protein
ORF01542	SAG1381	714	sulfatase
ORF01543	SAG1382	119	ribosomal protein L20
ORF01544	SAG1383	66	ribosomal protein L35
ORF01545	SAG1384	176	translation initiation factor IF-3
ORF01546	SAG1385	227	cytidylate kinase
ORF01547	SAG1386	174	conserved hypothetical protein
ORF01548	SAG1387	65	ferredoxin, 4Fe-4S
ORF01549	SAG1388	163	conserved hypothetical protein
ORF01550	SAG1389	406	peptidase t
ORF01551	SAG1390	544	polysaccharide biosynthesis protein, putative

Tabl 32: Conv rsion of ORF R f Nos. with SAG R f Nos.

ORF R f N .	SAGxxxx Ref No.	aa	Annotation
ORF01552	SAG1391	484	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelat ligase
ORF01553	SAG1392	264	iron compound ABC transporter, ATP-binding protein
ORF01554	SAG1393	310	iron compound ABC transporter, substrate-binding protein
ORF01555	SAG1394	341	iron compound ABC transporter, permease protein
ORF01556	SAG1395	333	iron compound ABC transporter, permease protein
ORF01557	SAG1396	217	conserved hypothetical protein
ORF01558	SAG1397	311	inorganic pyrophosphatase, manganese-dependent
ORF01559	SAG1398	262	pyruvate formate-lyase-activating enzyme
ORF01560	SAG1399	444	CBS domain protein
ORF01561	SAG1400	188	conserved hypothetical protein
ORF01563	SAG1401	311	conserved hypothetical protein TIGR01212
ORF01564	SAG1402	213	PAP2 family protein
ORF01565	SAG1403	194	membrane protein, putative
ORF01566	SAG1404	308	cell wall surface anchor family protein
ORF01567	SAG1405	294	sortase family protein
ORF01568	SAG1406	293	sortase family protein
ORF01569	SAG1407	705	cell wall surface anchor family protein
ORF01570	SAG1408	901	cell wall surface anchor family protein
ORF01571	SAG1409	326	transcriptional regulator, RofA family FRAMESHIFT
ORF01572	SAG1410	379	glycosyl transferase, group 1 family protein
ORF01573	SAG1411	282	exopolysaccharide biosynthesis protein, putative
ORF01574	SAG1412	474	exopolysaccharide biosynthesis protein, putative
ORF01575	SAG1413	454	hypothetical protein
ORF01576	SAG1414	308	glycosyl transferase, group 2 family protein
ORF01577	SAG1415	311	glycosyl transferase, group 2 family protein
ORF01578	SAG1416	352	dTDP-glucose 4,6-dehydratase, putative
ORF01579	SAG1417	240	4-diphosphocytidyl-2C-methyl-D-erythritol synthase, putative
ORF01580	SAG1418	259	licD protein, putative
ORF01581	SAG1419	577	hypothetical protein
ORF01582	SAG1420	117	conserved hypothetical protein
ORF01583	SAG1421	243	glycosyl transferase, group 2 family protein
ORF01584	SAG1422	313	glycosyl transferase, group 2 family protein
ORF01585	SAG1423	384	conserved hypothetical protein
ORF01586	SAG1424	284	dTDP-4-dehydrorhamnose reductase
ORF01587	SAG1425	113	conserved hypothetical protein
ORF01589	SAG1426	369	RNA polymerase sigma-70 factor
ORF01590	SAG1427	602	DNA primase
ORF01591	SAG1428	125	large conductance mechanosensitive channel protein
ORF01592	SAG1429	58	ribosomal protein S21
ORF01593	SAG1430	167	conserved hypothetical protein
ORF01594	SAG1431	268	amino acid ABC transporter, amino acid-binding protein
ORF01596	SAG1432	347	ammonium transporter family protein
ORF01597	SAG1433	375	conserved hypothetical protein
ORF01598	SAG1434	328	rhodanese family protein
ORF01599	SAG1435	101	conserved hypothetical protein
ORF01600	SAG1436	457	glycerol-3-phosphate transporter, putative

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01662	SAG1488	195	dephospho-CoA kinase
ORF01663	SAG1489	273	formamidopyrimidine-DNA glycosylase
ORF01665	SAG1490	282	transcriptional regulator, MutR family
ORF01666	SAG1491	530	hypothetical protein
ORF01667	SAG1492	58	hypothetical protein
ORF01668	SAG1493		hypothetical protein
ORF01670	SAG1494	32	hypothetical protein
ORF01672	SAG1495	81	protease, putative, POINT MUTATION
ORF01673	SAG1496	110	hypothetical protein
ORF01674	SAG1497	37	hypothetical protein
ORF01675	SAG1498	133	hypothetical protein
ORF01677	SAG1499	299	GTP-binding protein Era
ORF01678	SAG1500	132	diacylglycerol kinase
ORF01679	SAG1501	161	conserved hypothetical protein TIGR00043
ORF01680	SAG1502	268	tetracenomycin polyketide synthesis O-methyltransferase TcmP, putative
ORF01681	SAG1503	39	hypothetical protein
ORF01682	SAG1504	38	hypothetical protein
ORF01683	SAG1505	158	MutT/nudix family protein
ORF01684	SAG1506	267	hypothetical protein
ORF01685	SAG1507	345	PhoH family protein
ORF01686	SAG1508	590	67 kDa Myosin-crossreactive streptococcal antigen
ORF01687	SAG1509	71	conserved hypothetical protein
ORF01688	SAG1510	169	peptide methionine sulfoxide reductase
ORF01689	SAG1511	284	conserved hypothetical protein
ORF01690	SAG1512	185	ribosome recycling factor
ORF01691	SAG1513	242	uridylate kinase
ORF01692	SAG1514	226	peptide ABC transporter, ATP-binding protein
ORF01693	SAG1515	262	peptide ABC transporter, ATP-binding protein
ORF01694	SAG1516	255	peptide ABC transporter, permease protein
ORF01695	SAG1517	314	peptide ABC transporter, permease protein
ORF01696	SAG1518	525	peptide ABC transporter, peptide-binding protein
ORF01697	SAG1519	229	ribosomal protein L1
ORF01698	SAG1520	141	ribosomal protein L11
ORF01699	SAG1521	388	transposase, IS30 family, putative
ORF01700	SAG1522	460	transporter, major facilitator family
ORF01702	SAG1523	404	peptidase, M20/M25/M40 family
ORF01703	SAG1524	294	transcriptional regulator, LysR family
ORF01704	SAG1525	117	conserved hypothetical protein
ORF01705	SAG1526	178	IS861, transposase OrfA
ORF01706	SAG1527	277	IS861, transposase OrfB
ORF01707	SAG1528	571	chorismate binding enzyme
ORF01708	SAG1529	785	FtsK/SpoIIIE family protein
ORF01709	SAG1530	267	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
ORF01710	SAG1531	277	manganese ABC transporter, permease protein
ORF01711	SAG1532	238	manganese ABC transporter, ATP-binding protein
ORF01712	SAG1533	308	manganese ABC transporter, manganese-binding adhesion lipoprotein
ORF01713	SAG1534	215	iron-dependent transcriptional regulator
ORF01714	SAG1535	229	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase
ORF01715	SAG1536	89	conserved hypothetical protein
ORF01716	SAG1537	184	MutT/nudix family protein

Tabl 32: C nversion of ORF R f Nos. with SAG Ref N s.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF01718	SAG1538	459	UDP-N-acetylglucosamine pyrophosphorylase
ORF01719	SAG1539	31	hypothetical protein
ORF01720	SAG1540	137	conserved hypothetical protein
ORF01721	SAG1541	125	glyoxalase family protein
ORF01722	SAG1542	318	oxidoreductase, Gfo/ldh/MocA family
ORF01724	SAG1543		conserved hypothetical protein, FRAMESHIFT
ORF01725	SAG1544	232	gluconate 5-dehydrogenase, putative
ORF01726	SAG1545	78	conserved hypothetical protein
ORF01727	SAG1546	82	conserved hypothetical protein
ORF01729	SAG1547	166	acetyltransferase, GNAT family
ORF01730	SAG1548	422	glycosyl transferase, group 2 family protein
ORF01731	SAG1549	127	IS1381, transposase OrfA
ORF01732	SAG1550	129	IS1381, transposase OrfB
ORF01733	SAG1551	67	hypothetical protein
ORF01734	SAG1552	719	conserved hypothetical protein
ORF01735	SAG1553	477	hypothetical protein
ORF01736	SAG1554	225	hypothetical protein
ORF01737	SAG1555	231	hypothetical protein
ORF01738	SAG1556	445	branched-chain amino acid transport system II carrier protein
ORF01739	SAG1557	665	methionyl-tRNA synthetase
ORF01740	SAG1558	291	tellurite resistance protein TehB
ORF01741	SAG1559	231	membrane protein, putative
ORF01742	SAG1560	40	hypothetical protein
ORF01743	SAG1561	405	PTS system component, putative
ORF01744	SAG1562	280	conserved hypothetical protein
ORF01745	SAG1563	275	exodeoxyribonuclease
ORF01746	SAG1564	118	conserved hypothetical protein
ORF01747	SAG1565	158	methylated-DNA-protein-cysteine S-methyltransferase
ORF01748	SAG1566	393	D-isomer specific 2-hydroxyacid dehydrogenase family protein
ORF01749	SAG1567	182	acetyltransferase, GNAT family
ORF01750	SAG1568		phosphoserine aminotransferase FRAMESHIFT
ORF01752	SAG1569	211	copper homeostasis protein CutC, putative
ORF01753	SAG1570	34	conserved hypothetical protein
ORF01754	SAG1571	53	hypothetical protein
ORF01755	SAG1572	287	tetrapyrrole methylase family protein
ORF01756	SAG1573	108	conserved hypothetical protein
ORF01758	SAG1574	287	DNA polymerase III, delta prime subunit, putative
ORF01759	SAG1575	211	thymidylate kinase
ORF01761	SAG1576	267	transposase, IS30 family, putative, truncation
ORF01763	SAG1577	219	AcuB family protein
ORF01764	SAG1578	236	branched-chain amino acid ABC transporter, ATP-binding protein
ORF01765	SAG1579	254	branched-chain amino acid ABC transporter, ATP-binding protein
ORF01766	SAG1580	317	branched-chain amino acid ABC transporter, permease protein
ORF01767	SAG1581	289	branched-chain amino acid ABC transporter, permease protein
ORF01769	SAG1582	388	branched-chain amino acid ABC transporter, amino acid-binding protein
ORF01770	SAG1583	81	conserved hypothetical protein
ORF01772	SAG1584	377	IS1548, transposase

Tabl 32: Conv rsion of ORF Ref N s. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Ann tati n
ORF01773	SAG1585	196	ATP-dependent Clp protease, proteolytic subunit ClpP
ORF01774	SAG1586	209	uracil phosphoribosyltransferase
ORF01775	SAG1587	389	aminotransferase, class I
ORF01777	SAG1588	182	RNA methyltransferase, TrmH family, group 2
ORF01778	SAG1589	450	amino acid permease, putative
ORF01779	SAG1590	449	potassium uptake protein, Trk family
ORF01780	SAG1591	475	cation uptake protein, Trk family
ORF01781	SAG1592	83	conserved hypothetical protein TIGR00278
ORF01782	SAG1593	240	ribosomal large subunit pseudouridine synthase B
ORF01783	SAG1594	194	conserved hypothetical protein TIGR00281
ORF01784	SAG1595	235	Uncharacterized ACR, COG1354
ORF01785	SAG1596	246	integrase/recombinase, phage integrase family
ORF01786	SAG1597	157	CBS domain protein
ORF01787	SAG1598	173	conserved hypothetical protein
ORF01788	SAG1599	324	HAM1 protein
ORF01789	SAG1600	264	glutamate racemase
ORF01790	SAG1601	79	conserved hypothetical protein
ORF01791	SAG1602	180	membrane protein, putative
ORF01792	SAG1603	173	transcriptional regulator, biotin repressor family
ORF01793	SAG1604	229	membrane protein, putative
ORF01794	SAG1605	167	conserved hypothetical protein
ORF01795	SAG1606	247	RNA methyltransferase, TrmH family
ORF01796	SAG1607	92	acylphosphatase
ORF01797	SAG1608	310	membrane protein, putative
ORF01799	SAG1609	221	amino acid ABC transporter, permease protein
ORF01800	SAG1610	285	amino acid ABC transporter, substrate-binding protein
ORF01801	SAG1611	486	amidase family protein
ORF01802	SAG1612	160	transcription elongation factor GreA
ORF01803	SAG1613	600	Uncharacterized BCR, YceG family COG1559, putative
ORF01804	SAG1614	167	acetyltransferase, GNAT family
ORF01805	SAG1615	443	UDP-N-acetylmuramate-alanine ligase
ORF01806	SAG1616	205	conserved hypothetical protein
ORF01807	SAG1617	32	hypothetical protein
ORF01808	SAG1618	1032	Snf2 family protein
ORF01810	SAG1619	377	IS1548, transposase
ORF01811	SAG1620	436	phosphoglycerate dehydrogenase-related protein
ORF01812	SAG1621	300	primosomal protein DnaI
ORF01813	SAG1622	391	conserved hypothetical protein
ORF01814	SAG1623	159	conserved hypothetical protein TIGR00244
ORF01815	SAG1624	501	sensor histidine kinase CsrS
ORF01816	SAG1625	229	DNA-binding response regulator CsrR
ORF01817	SAG1626	177	conserved hypothetical protein
ORF01818	SAG1627	296	heat shock protein HtpX
ORF01820	SAG1628	184	IemA protein
ORF01821	SAG1629	237	glucose-inhibited division protein B
ORF01822	SAG1630	459	sodium transport family protein
ORF01823	SAG1631	223	potassium uptake protein, Trk family, putative
ORF01824	SAG1632	276	cobalt transport family protein
ORF01825	SAG1633	558	ABC transporter, ATP-binding protein
ORF01826	SAG1634	212	conserved hypothetical protein
ORF01827	SAG1635	402	sodium:dicarboxylate symporter family protein

Tabl 32: Conversion f ORF Ref Nos. with SAG R f N s.

ORF Ref N .	SAGxxxx Ref N .	aa	Annotation
ORF01828	SAG1636	455	branched-chain amino acid transport system II carrier protein
ORF01829	SAG1637	351	alcohol dehydrogenase, zinc-containing
ORF01830	SAG1638	230	ABC transporter, permease protein
ORF01831	SAG1639	356	ABC transporter, ATP-binding protein
ORF01832	SAG1640	458	peptidase, M20/M25/M40 family
ORF01833	SAG1641	274	lipoprotein, putative
ORF01834	SAG1642	277	ABC transporter, substrate-binding protein
ORF01835	SAG1643	229	glutamine amidotransferase, class I
ORF01836	SAG1644	37	hypothetical protein
ORF01837	SAG1645	238	conserved hypothetical protein TIGR01033
ORF01838	SAG1646	32	hypothetical protein
ORF01839	SAG1647	328	dihydroxyacetone kinase family protein
ORF01840	SAG1648	178	transcriptional regulator, TetR family, putative
ORF01842	SAG1649	37	hypothetical protein
ORF01843	SAG1650	329	dihydroxyacetone kinase family protein
ORF01844	SAG1651	192	dihydroxyacetone kinase family protein
ORF01845	SAG1652	124	conserved hypothetical protein
ORF01846	SAG1653	237	glycerol uptake facilitator protein
ORF01847	SAG1654	134	conserved hypothetical protein
ORF01848	SAG1655	237	transcriptional regulator, MerR family
ORF01849	SAG1656	369	conserved hypothetical protein
ORF01850	SAG1657	83	hypothetical protein
ORF01851	SAG1658	244	conserved hypothetical protein
ORF01852	SAG1659	118	iojap-related protein
ORF01853	SAG1660	173	isochorismatase family protein
ORF01854	SAG1661	195	conserved hypothetical protein TIGR00488
ORF01855	SAG1662	210	conserved hypothetical protein TIGR00482
ORF01856	SAG1663	105	conserved hypothetical protein TIGR00253
ORF01857	SAG1664	372	GTP-binding protein
ORF01858	SAG1665	177	hydrolase, haloacid dehalogenase-like family
ORF01859	SAG1666	295	membrane protein
ORF01860	SAG1667	480	glutamyl-tRNA(Gln) amidotransferase, B subunit
ORF01861	SAG1668	488	glutamyl-tRNA(Gln) amidotransferase, A subunit
ORF01862	SAG1669	100	glutamyl-tRNA(Gln) amidotransferase, C subunit
ORF01863	SAG1670	881	pyruvate phosphate dikinase
ORF01864	SAG1671	276	conserved hypothetical protein
ORF01865	SAG1672	170	CBS domain protein
ORF01866	SAG1673	377	3-hydroxyacyl-CoA dehydrogenase family protein
ORF01867	SAG1674	182	isochorismatase family protein
ORF01869	SAG1675	261	transcriptional regulator CodY, putative
ORF01870	SAG1676	403	aminotransferase, class I
ORF01871	SAG1677	137	universal stress protein family FRAMESHIFT
ORF01872	SAG1678	460	hydrolase, haloacid dehalogenase-like family
ORF01873	SAG1679	320	asparaginase family protein
ORF01874	SAG1680	292	shikimate 5-dehydrogenase
ORF01875	SAG1681	304	oxidoreductase, aldo/keto reductase family
ORF01876	SAG1682	671	ATP-dependent DNA helicase RecG
ORF01877	SAG1683	512	immunogenic secreted protein, putative
ORF01878	SAG1684	366	alanin racemase
ORF01879	SAG1685	119	holo-(acyl-carrier-protein) synthase
ORF01880	SAG1686	335	phospho-2-dehydro-3-deoxyheptonate aldolase
ORF01881	SAG1687	842	preprotein translocase, SecA subunit
ORF01882	SAG1688	315	mannose-6-phosphate isomerase, class I
ORF01883	SAG1689	293	fructokinase

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref N .	aa	Annotati n
ORF01885	SAG1690	639	PTS system IIABC components
ORF01886	SAG1691	479	sucrose-6-phosphate hydrolase
ORF01887	SAG1692	320	sucrose operon repressor ScrR
ORF01888	SAG1693	144	N utilization substance protein B
ORF01889	SAG1694	129	conserved hypothetical protein
ORF01890	SAG1695	186	translation elongation factor P
ORF01892	SAG1696	38	hypothetical protein
ORF01893	SAG1697	48	hypothetical protein
ORF01894	SAG1698	99	conserved hypothetical protein
ORF01895	SAG1699	30	hypothetical protein
ORF01896	SAG1700	76	hypothetical protein
ORF01897	SAG1701	56	hypothetical protein
ORF01898	SAG1702	41	hypothetical protein
ORF01899	SAG1703	54	hypothetical protein
ORF01900	SAG1704	150	cytidine/deoxycytidylate deaminase family protein
ORF01902	SAG1705		peptidase, M24 family POINT MUTATION
ORF01903	SAG1706	238	conserved hypothetical protein
ORF01904	SAG1707	499	drug resistance transporter, EmrB/QacA family
ORF01905	SAG1708	38	hypothetical protein
ORF01906	SAG1709	942	excinuclease ABC, A subunit
ORF01907	SAG1710	223	conserved hypothetical protein
ORF01908	SAG1711	314	magnesium transporter, CorA family
ORF01909	SAG1712	79	ribosomal protein S18
ORF01910	SAG1713	163	single-strand binding protein
ORF01911	SAG1714	95	ribosomal protein S6
ORF01912	SAG1715	374	A/G-specific adenine glycosylase
ORF01913	SAG1716	197	transcriptional regulator, Cro/C1 family
ORF01914	SAG1717	104	thioredoxin
ORF01915	SAG1718	166	PAP2 family protein
ORF01916	SAG1719	779	MutS2 family protein
ORF01917	SAG1720	180	conserved hypothetical protein
ORF01918	SAG1721	103	conserved hypothetical protein
ORF01919	SAG1722	297	ribonuclease HIII
ORF01920	SAG1723	197	signal peptidase I
ORF01921	SAG1724	806	helicase, putative
ORF01922	SAG1725	160	conserved hypothetical protein
ORF01923	SAG1726	364	DNA-damage inducible protein P
ORF01924	SAG1727	770	formate acetyltransferase
ORF01925	SAG1728	124	FMN-binding protein
ORF01926	SAG1729	309	conserved hypothetical protein
ORF01927	SAG1730	251	proteinase, putative, degenerate, FRAMESHIFT
ORF01928	SAG1731	298	membrane protein, putative
ORF01929	SAG1732	282	glycerol uptake facilitator protein, putative
ORF01930	SAG1733	150	universal stress protein family
ORF01931	SAG1734	400	transporter, putative
ORF01932	SAG1735	219	transcriptional regulator, Crp/Fnr family
ORF01933	SAG1736	761	X-pro dipeptidyl-peptidase
ORF01934	SAG1737	119	hypothetical protein
ORF01936	SAG1738	326	polyprenyl synthetase family protein
ORF01937	SAG1739	582	ABC transporter, ATP-binding protein CydC
ORF01938	SAG1740	572	ABC transporter, ATP-binding protein CydD
ORF01939	SAG1741	339	cytochrome d ubiquinol oxidase, subunit II
ORF01940	SAG1742	475	cytochrome d oxidase, subunit I
ORF01941	SAG1743	402	pyridine nucleotide-disulphide oxidoreductase family protein

Tabl 32: Conv rsion of ORF R f N s. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotati n
ORF01942	SAG1744	299	prenyltransferase, UbiA family
ORF01943	SAG1745	148	hypothetical protein
ORF01944	SAG1746	35	hypothetical protein
ORF01945	SAG1747	99	conserved hypothetical protein TIGR00103
ORF01946	SAG1748	396	cyclopropane-fatty-acyl-phospholipid synthase
ORF01947	SAG1749	241	transcriptional regulator, merR family
ORF01948	SAG1750	195	exonuclease
ORF01949	SAG1751	178	conserved hypothetical protein
ORF01950	SAG1752	375	conserved hypothetical protein TIGR00275
ORF01951	SAG1753	260	conserved hypothetical protein
ORF01952	SAG1754	89	ribosomal protein S14
ORF01953	SAG1755	38	hypothetical protein
ORF01954	SAG1756	341	conserved hypothetical protein
ORF01957	SAG1757	336	O-sialoglycoprotein endopeptidase family protein
ORF01958	SAG1758	135	ribosomal-protein-alanine acetyltransferase, putative
ORF01960	SAG1759	230	glycoprotease family protein, putative
ORF01961	SAG1760	76	conserved hypothetical protein
ORF01962	SAG1761	559	metallo-beta-lactamase superfamily protein
ORF01963	SAG1762	169	conserved hypothetical protein
ORF01964	SAG1763	448	glutamine synthetase, type I
ORF01965	SAG1764	123	transcriptional regulator GinR
ORF01967	SAG1765	179	conserved hypothetical protein
ORF01969	SAG1766	398	phosphoglycerate kinase
ORF01970	SAG1767	289	acid phosphatase
ORF01971	SAG1768	336	glyceraldehyde 3-phosphate dehydrogenase
ORF01972	SAG1769	692	translation elongation factor G
ORF01973	SAG1770	156	ribosomal protein S7
ORF01974	SAG1771	137	ribosomal protein S12
ORF01975	SAG1772	270	pur operon repressor
ORF01976	SAG1773	313	HD domain protein
ORF01977	SAG1774	424	conserved hypothetical protein
ORF01978	SAG1775	210	conserved hypothetical protein
ORF01979	SAG1776	220	ribulose-phosphate 3-epimerase
ORF01980	SAG1777	290	conserved hypothetical protein TIGR00157
ORF01981	SAG1778	283	rRNA (guanine-N1-)-methyltransferase, putative
ORF01983	SAG1779	290	dimethyladenosine transferase
ORF01984	SAG1780	163	hypothetical protein
ORF01985	SAG1781	186	primase-related protein
ORF01987	SAG1782	260	deoxyribonuclease, TatD family
ORF01988	SAG1783	90	hypothetical protein
ORF01989	SAG1784	130	hypothetical protein
ORF01990	SAG1785	430	hypothetical protein
ORF01991	SAG1786	130	hypothetical protein
ORF01992	SAG1787	420	ditD protein
ORF01993	SAG1788	79	D-alanyl carrier protein
ORF01994	SAG1789	421	ditB protein
ORF01996	SAG1790	511	D-alanine-activating enzyme
ORF01997	SAG1791	395	sensor histidine kinase
ORF01998	SAG1792	224	DNA-binding response regulator
ORF01999	SAG1793	44	ribosomal protein L34
ORF02000	SAG1794	451	membrane protein, putative
ORF02001	SAG1795	388	transposase, IS30 family, putative
ORF02002	SAG1796	575	amino acid ABC transporter, permease protein
ORF02004	SAG1797	407	amino acid ABC transporter, ATP-binding protein

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF02005	SAG1798	39	hypothetical protein
ORF02006	SAG1799	792	xylulose-5-phosphate/fructose-6-phosphate phosphoketolase
ORF02007	SAG1800	363	conserved hypothetical protein
ORF02008	SAG1801	559	transcriptional antiterminator, BglG family
ORF02009	SAG1802	253	conserved hypothetical protein
ORF02010	SAG1803	505	carbohydrate kinase, FGGY family
ORF02011	SAG1804	329	hypothetical protein
ORF02012	SAG1805	483	PTS system component, putative
ORF02015	SAG1806	318	glyoxylate reductase, NADH-dependent
ORF02016	SAG1807	339	hypothetical protein
ORF02017	SAG1808	327	sugar binding transcriptional regulator, LacI family
ORF02018	SAG1809	215	transaldolase family protein
ORF02019	SAG1810	238	carbohydrate isomerase, AraD/FucA family
ORF02020	SAG1811	287	hexulose-6-phosphate isomerase, putative
ORF02021	SAG1812	221	hexulose-6-phosphate synthase, putative
ORF02022	SAG1813	161	PTS system, IIA component
ORF02023	SAG1814	92	PTS system, IIB component
ORF02024	SAG1815	479	transport protein SgaT, putative
ORF02025	SAG1816	205	hypothetical protein
ORF02026	SAG1817	157	hypothetical protein
ORF02027	SAG1818	430	adenylosuccinate synthetase
ORF02028	SAG1819	340	perfringolysin O regulator protein
ORF02029	SAG1820	224	conserved hypothetical protein
ORF02030	SAG1821	750	glutamate-cysteine ligase-related protein
ORF02031	SAG1822	272	conserved hypothetical protein
ORF02032	SAG1823	418	conserved hypothetical protein
ORF02033	SAG1824	291	chaperonin, 33 kDa
ORF02034	SAG1825	325	NifR3/Smm1 family protein
ORF02035	SAG1826	213	deoxynucleoside kinase family protein
ORF02036	SAG1827	163	phosphinothricin N-acetyltransferase
ORF02037	SAG1828	815	ATP-dependent Clp protease, ATP-binding subunit
ORF02038	SAG1829	154	transcriptional regulator CtsR
ORF02039	SAG1830	153	conserved hypothetical protein
ORF02040	SAG1831	346	translation elongation factor Ts
ORF02041	SAG1832	256	ribosomal protein S2
ORF02042	SAG1833	186	alkyl hydroperoxide reductase, subunit C
ORF02043	SAG1834	510	alkyl hydroperoxide reductase, subunit F
ORF02044	SAG1835	134	conserved hypothetical protein
ORF02045	SAG1836	61	conserved hypothetical protein
ORF02046	SAG1837	468	lysine, putative
ORF02047	SAG1838	109	holin, putative
ORF02048	SAG1839	136	conserved hypothetical protein
ORF02049	SAG1840	112	hypothetical protein
ORF02050	SAG1841	76	conserved domain protein
ORF02051	SAG1842	1224	PblB, putative
ORF02053	SAG1843	240	conserved hypothetical protein
ORF02056	SAG1844	911	conserved hypothetical protein
ORF02057	SAG1845	42	hypothetical protein
ORF02058	SAG1846	158	hypothetical protein
ORF02059	SAG1847	227	conserved hypothetical protein
ORF02060	SAG1848	114	conserved hypothetical protein
ORF02061	SAG1849	115	hypothetical protein

Table 32: Comparison of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref N .	aa	Annotation
ORF02062	SAG1850	101	hypothetical protein
ORF02063	SAG1851	111	conserved domain protein
ORF02064	SAG1852	420	conserved domain protein
ORF02066	SAG1853	180	protease, putative
ORF02067	SAG1854	380	conserved hypothetical protein
ORF02068	SAG1855	570	terminase large subunit, putative
ORF02069	SAG1856	161	hypothetical protein
ORF02070	SAG1858	95	hypothetical protein
ORF02071	SAG1859	180	site-specific recombinase, phage integrase family
ORF02072	SAG1860	154	conserved hypothetical protein
ORF02073	SAG1861	119	transcriptional regulator, Cro/C1 family
ORF02075	SAG1862	86	hypothetical protein
ORF02076	SAG1863	138	single-strand binding protein
ORF02077	SAG1864	68	hypothetical protein
ORF02078	SAG1865	74	conserved hypothetical protein
ORF02079	SAG1866	109	conserved hypothetical protein
ORF02080	SAG1867	163	conserved hypothetical protein
ORF02081	SAG1868	134	hypothetical protein
ORF02082	SAG1869	437	type II DNA modification methyltransferase, putative
ORF02083	SAG1870	273	DNA replication protein DnaC, putative
ORF02084	SAG1871	248	conserved hypothetical protein
ORF02085	SAG1872	200	hypothetical protein
ORF02086	SAG1873	443	replicative DNA helicase
ORF02087	SAG1874	87	hypothetical protein
ORF02088	SAG1875	94	conserved hypothetical protein
ORF02089	SAG1876	176	HNH endonuclease family protein
ORF02090	SAG1877	236	antirepressor protein, putative
ORF02091	SAG1878	102	conserved domain protein
ORF02092	SAG1879	156	hypothetical protein
ORF02093	SAG1880	54	hypothetical protein
ORF02094	SAG1881	51	hypothetical protein
ORF02095	SAG1882	120	repressor protein, putative
ORF02097	SAG1884	134	hypothetical protein
ORF02098	SAG1885	356	site-specific recombinase, phage integrase family
ORF02100	SAG1886	32	hypothetical protein
ORF02101	SAG1887	689	Na ⁺ /H ⁺ exchanger family protein
ORF02102	SAG1888	78	hypothetical protein
ORF02103	SAG1889	317	microcin immunity protein MccF, putative
ORF02104	SAG1890	631	endopeptidase O
ORF02105	SAG1891	327	oxidoreductase, Gfo/ldh/MocA family
ORF02107	SAG1892	358	membrane protein, putative
ORF02108	SAG1893	59	hypothetical protein
ORF02109	SAG1894	214	Cyclic nucleotide-binding domain protein
ORF02110	SAG1895	204	polypeptide deformylase
ORF02111	SAG1896	333	sugar binding transcriptional regulator RegR
ORF02112	SAG1897	634	conserved hypothetical protein
ORF02113	SAG1898	271	PTS system, IID component
ORF02114	SAG1899	288	PTS system, IIC component
ORF02115	SAG1900	164	PTS system, IIB component
ORF02116	SAG1901	398	glucuronyl hydrolase
ORF02118	SAG1902	144	PTS system, IIA component
ORF02119	SAG1903	34	hypothetical protein

Tabl 32: Conversion of ORF R f N s. with SAG Ref N s.

ORF Ref N .	SAGxxxx Ref No.	aa	Annotati n
ORF02120	SAG1904	270	oxidoreductase, short-chain dehydrogenase/reductase family
ORF02121	SAG1905	212	conserved hypothetical protein
ORF02122	SAG1906	335	carbohydrate kinase, PfkB family
ORF02123	SAG1907	212	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase
ORF02124	SAG1908	499	hypothetical protein
ORF02125	SAG1909	204	nitroreductase family protein
ORF02126	SAG1910	141	transcriptional regulator, MarR family
ORF02127	SAG1911	1468	DNA polymerase III, alpha subunit, Gram-positive type
ORF02128	SAG1912	194	N-acetylmuramoyl-L-alanine amidase, family 4 protein
ORF02129	SAG1913	617	prolyl-tRNA synthetase
ORF02130	SAG1914	419	membrane-associated zinc metalloprotease, putative
ORF02131	SAG1915	264	phosphatidate cytidylyltransferase
ORF02132	SAG1916	250	undecaprenyl diphosphate synthase
ORF02133	SAG1917	113	preprotein translocase, YajC subunit
ORF02134	SAG1918	114	conserved hypothetical protein
ORF02135	SAG1919	387	malate oxidoreductase
ORF02136	SAG1920	445	citrate carrier protein, CCS family
ORF02137	SAG1921	508	sensor histidine kinase family protein
ORF02138	SAG1922	229	response regulator
ORF02139	SAG1923	331	UDP-glucose 4-epimerase
ORF02140	SAG1924	535	glucan 1,6-alpha-glucosidase
ORF02141	SAG1925	377	sugar ABC transporter, ATP-binding protein
ORF02142	SAG1926	283	helix-turn-helix domain protein, fis-type
ORF02143	SAG1927	298	lacX protein
ORF02144	SAG1928	325	tagatose 1,6-diphosphate aldolase
ORF02145	SAG1929	310	tagatose-6-phosphate kinase
ORF02146	SAG1930	171	galactose-6-phosphate isomerase, LacB subunit
ORF02147	SAG1931	141	galactose-6-phosphate isomerase, LacA subunit
ORF02148	SAG1932	816	neuraminidase
ORF02149	SAG1933	482	PTS system, IIC component, putative
ORF02150	SAG1934	101	PTS system, IIB component, putative
ORF02152	SAG1935	157	PTS system, IIA component, putative
ORF02153	SAG1936	258	lactose phosphotransferase system repressor
ORF02156	SAG1937		streptococcal histidine triad family protein, degenerate, FRAMESHIFT
ORF02157	SAG1938	307	adhesion lipoprotein, putative
ORF02158	SAG1939	147	conserved hypothetical protein TIGR00256
ORF02159	SAG1940	738	GTP pyrophosphokinase
ORF02160	SAG1941	800	2',3'-cyclic-nucleotide 2'-phosphodiesterase
ORF02161	SAG1942	151	nrdf protein, putative
ORF02162	SAG1943	345	conserved hypothetical protein
ORF02163	SAG1944	165	conserved hypothetical protein
ORF02164	SAG1945	345	iron ABC transporter, iron-binding protein
ORF02165	SAG1946	257	DNA-binding response regulator
ORF02166	SAG1947	549	conserved hypothetical protein
ORF02167	SAG1948	275	PTS system, IID component
ORF02168	SAG1949	269	PTS system, IIC component
ORF02169	SAG1950	163	PTS system, IIB component
ORF02170	SAG1951	141	PTS system, IIA component, putative
ORF02171	SAG1952	353	membrane protein, putative
ORF02172	SAG1953	60	hypothetical pr tein

Tabl 32: Conv rsion of ORF Ref N s. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Ann tation
ORF02173	SAG1954	384	hypothetical protein
ORF02174	SAG1955	282	ABC transporter, ATP-binding protein
ORF02175	SAG1956	96	conserved domain protein
ORF02176	SAG1957	250	response regulator
ORF02177	SAG1958	276	conserved hypothetical protein
ORF02178	SAG1959	727	PTS system, IIABC components
ORF02179	SAG1960	551	sensor histidine kinase
ORF02180	SAG1961	225	phosphate regulon response regulator PhoB
ORF02181	SAG1962	218	phosphate transport system regulatory protein PhoU, putative
ORF02182	SAG1963	253	phosphate ABC transporter, ATP-binding protein
ORF02183	SAG1964	292	phosphate ABC transporter, permease protein
ORF02184	SAG1965	281	phosphate ABC transporter, permease protein
ORF02186	SAG1966	293	hemolysin precursor, putative
ORF02187	SAG1967	195	hypothetical protein
ORF02188	SAG1968	246	conserved hypothetical protein TIGR00046
ORF02189	SAG1969	317	ribosomal protein L11 methyltransferase
ORF02190	SAG1970	102	conserved hypothetical protein
ORF02191	SAG1971	41	hypothetical protein
ORF02192	SAG1972	238	transcriptional regulator, MerR family
ORF02194	SAG1973	156	acetyltransferase, GNAT family
ORF02195	SAG1974	152	MutT/nudix family protein
ORF02196	SAG1975	47	hypothetical protein
ORF02197	SAG1976	156	conserved hypothetical protein
ORF02198	SAG1977	163	acetyltransferase, GNAT family
ORF02199	SAG1978	422	ATPase, AAA family
ORF02201	SAG1979	253	hypothetical protein
ORF02202	SAG1980	300	ABC transporter, ATP-binding protein
ORF02203	SAG1981	68	hypothetical protein
ORF02205	SAG1982	359	transcriptional regulator, Cro/C1 family
ORF02206	SAG1983	105	conserved hypothetical protein
ORF02207	SAG1984	188	conserved hypothetical protein TIGR00730
ORF02208	SAG1985	51	hypothetical protein
ORF02209	SAG1986	375	integrase, phage family, putative
ORF02210	SAG1987	61	conserved hypothetical protein
ORF02211	SAG1988	342	conserved hypothetical protein
ORF02212	SAG1989	139	hypothetical protein
ORF02213	SAG1990	127	hypothetical protein
ORF02214	SAG1991	204	transcriptional regulator, Cro/C1 family
ORF02215	SAG1992	518	conserved hypothetical protein
ORF02216	SAG1993	373	site-specific recombinase, phage integrase family
ORF02217	SAG1994	108	conserved hypothetical protein
ORF02219	SAG1995	210	hypothetical protein
ORF02221	SAG1996	263	cell wall anchor protein-related protein
ORF02223	SAG1997	182	hypothetical protein
ORF02224	SAG1998	457	hypothetical protein
ORF02225	SAG1999	47	hypothetical protein
ORF02226	SAG2000	666	membrane protein, putative
ORF02227	SAG2001	756	conjugal transfer protein, interruption-C
ORF02228	SAG2002	129	IS1381, transposase OrfB
ORF02229	SAG2003	127	IS1381, transposase OrfA
ORF02230	SAG2005	138	conserved hypothetical protein
ORF02231	SAG2006	88	conserved hypothetical protein
ORF02232	SAG2007	317	conserved hypothetical protein

Tabl 32: Conv rsion f ORF R f Nos. with SAG R f Nos.

ORF R f No.	SAGxxxx Ref No.	aa	Annotation
ORF02233	SAG2008	84	conserved hypothetical protein
ORF02234	SAG2009	88	conserved hypothetical protein
ORF02235	SAG2010	157	hypothetical protein
ORF02236	SAG2011	160	conserved hypothetical protein
ORF02237	SAG2012	90	hypothetical protein
ORF02238	SAG2013	189	hypothetical protein
ORF02239	SAG2014	449	hypothetical protein
ORF02240	SAG2015	99	transcriptional regulator, Cro/Ci family
ORF02241	SAG2016	125	hypothetical protein
ORF02242	SAG2017	429	transcriptional regulator, Cro/Ci family
ORF02243	SAG2018	553	FtsK/SpoIIIE family protein
ORF02244	SAG2019	153	hypothetical protein
ORF02245	SAG2020	98	hypothetical protein
ORF02246	SAG2021	826	cell wall surface anchor family protein
ORF02247	SAG2022	417	transposase, ISL3 family
ORF02249	SAG2023	546	mercuric reductase
ORF02250	SAG2024	130	mercuric resistance operon regulatory protein MerR
ORF02251	SAG2025	522	Mn2+/Fe2+ transporter, NRAMP family
ORF02252	SAG2026	240	membrane protein, putative
ORF02253	SAG2027	205	ABC transporter, ATP-binding protein
ORF02254	SAG2028	36	conserved hypothetical protein
ORF02255	SAG2029	284	streptomycin resistance protein
ORF02257	SAG2030	130	hypothetical protein
ORF02258	SAG2031	202	hypothetical protein
ORF02259	SAG2032	111	conserved hypothetical protein
ORF02260	SAG2033	162	acetyltransferase, GNAT family
ORF02261	SAG2034	247	membrane protein, putative
ORF02262	SAG2035	300	ABC transporter, ATP-binding protein
ORF02263	SAG2036	68	hypothetical protein
ORF02264	SAG2037	358	transcriptional regulator, Cro/Ci family
ORF02265	SAG2038	204	PAP2 family protein
ORF02266	SAG2039	98	conserved hypothetical protein
ORF02267	SAG2040	186	conserved hypothetical protein TIGR00730
ORF02268	SAG2041	287	protease, putative
ORF02269	SAG2042	100	rhodanese family protein
ORF02270	SAG2043	255	cAMP factor
ORF02271	SAG2044	62	hypothetical protein
ORF02272	SAG2045	179	DNA topology modulation protein FlaR, putative
ORF02273	SAG2046	361	glycerol dehydrogenase, putative
ORF02274	SAG2047	235	conserved hypothetical protein
ORF02275	SAG2048	614	5-methyltetrahydrofolate--homocysteine methyltransferase, putative
ORF02276	SAG2049	745	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
ORF02277	SAG2050	107	conserved hypothetical protein
ORF02278	SAG2051	230	branched-chain amino acid transport protein AzlC, putative
ORF02279	SAG2052	41	hypothetical protein
ORF02280	SAG2053	1570	serine protease, subtilase family, putative
ORF02281	SAG2054	228	DNA-binding response regulator
ORF02282	SAG2055	462	sensor histidine kinase
ORF02283	SAG2056	202	chromosome assembly-relat d protein
ORF02285	SAG2057	833	leucyl-tRNA synthetase
ORF02286	SAG2058	415	major facilitator family protein

Table 32: Comparison of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF02287	SAG2059	281	conserved hypothetical protein
ORF02288	SAG2060	398	glycosyl transferase, family 8
ORF02289	SAG2061	401	glycosyl transferase, family 8
ORF02290	SAG2062	179	transcription antitermination protein NusG
ORF02291	SAG2063	630	pathogenicity protein, putative
ORF02292	SAG2064	57	preprotein translocase, SecE subunit, putative
ORF02293	SAG2066	773	penicillin-binding protein 2A
ORF02294	SAG2067	294	ribosomal large subunit pseudouridine synthase, RluD subfamily
ORF02295	SAG2068	546	Lyme disease proteins of unknown function, putative
ORF02296	SAG2069	403	phosphopentomutase
ORF02297	SAG2070	223	deoxyribose-phosphate aldolase
ORF02298	SAG2071	400	Na ⁺ dependent nucleoside transporter
ORF02300	SAG2072	259	uridine phosphorylase
ORF02301	SAG2073	245	transcriptional regulator, GntR family
ORF02302	SAG2074	540	60 kda chaperonin
ORF02303	SAG2075	94	chaperonin, 10 kDa
ORF02305	SAG2076	267	ABC transporter, ATP-binding protein
ORF02306	SAG2077	298	ABC transporter, permease protein
ORF02307	SAG2078	320	lipoprotein, putative
ORF02308	SAG2079	265	hydrolase, haloacid dehalogenase-like family
ORF02309	SAG2080	286	glyoxalase family protein
ORF02310	SAG2081	243	conserved hypothetical protein
ORF02311	SAG2082	205	anaerobic ribonucleoside-triphosphate reductase activating protein
ORF02312	SAG2083	163	acetyltransferase, GNAT family
ORF02313	SAG2084	310	virulence factor MviM, putative
ORF02314	SAG2085	47	conserved hypothetical protein
ORF02315	SAG2086	723	anaerobic ribonucleoside-triphosphate reductase
ORF02316	SAG2087	495	conserved hypothetical protein
ORF02317	SAG2088	40	hypothetical protein
ORF02318	SAG2089	105	conserved hypothetical protein
ORF02319	SAG2090	136	conserved hypothetical protein TIGR00250
ORF02320	SAG2091	88	conserved hypothetical protein
ORF02321	SAG2092	132	conserved hypothetical protein
ORF02322	SAG2093	379	recA protein
ORF02323	SAG2094		competence/damage-inducible protein ClnA FRAMESHIFT
ORF02325	SAG2095	183	DNA-3-methyladenine glycosylase I
ORF02327	SAG2096	196	Holliday junction DNA helicase RuvA
ORF02328	SAG2097	418	transporter, putative
ORF02329	SAG2098	659	DNA mismatch repair protein HexB
ORF02330	SAG2099	33	hypothetical protein
ORF02331	SAG2100	67	cold shock protein, CSD family
ORF02332	SAG2101	858	DNA mismatch repair protein HexA
ORF02333	SAG2102	145	arginine repressor ArgR, putative
ORF02334	SAG2103	563	arginyl-tRNA synthetase
ORF02335	SAG2104	102	conserved hypothetical protein
ORF02337	SAG2105	290	conserved hypothetical protein
ORF02338	SAG2106	314	conserved hypothetical protein
ORF02339	SAG2107	583	aspartyl-tRNA synthetase
ORF02340	SAG2108	426	histidyl-tRNA synthetase
ORF02341	SAG2109	60	ribosomal protein L32

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF02342	SAG2110	49	ribosomal protein L33
ORF02343	SAG2111	173	conserved hypothetical protein
ORF02344	SAG2112	494	site-specific recombinase, phage integrase family
ORF02345	SAG2113	82	conserved hypothetical protein
ORF02346	SAG2114	342	conserved hypothetical protein
ORF02347	SAG2115	143	hypothetical protein
ORF02349	SAG2116	151	conserved hypothetical protein
ORF02350	SAG2117	71	hypothetical protein
ORF02351	SAG2118	308	transcriptional regulator, Cro/Ci family
ORF02352	SAG2119	373	conserved domain protein
ORF02355	SAG2120	56	hypothetical protein
ORF02356	SAG2121	176	hypothetical protein
ORF02357	SAG2122	223	DNA-binding response regulator
ORF02358	SAG2123	454	sensor histidine kinase
ORF02359	SAG2124	517	membrane protein, putative
ORF02360	SAG2125	308	carbamate kinase
ORF02361	SAG2126	332	ornithine carbamoyltransferase
ORF02362	SAG2127	431	sensor histidine kinase
ORF02363	SAG2128	277	response regulator
ORF02364	SAG2129	240	amino acid ABC transporter, ATP-binding protein
ORF02365	SAG2130	504	amino acid ABC transporter, permease and amino acid binding protein
ORF02367	SAG2131	847	membrane protein, putative
ORF02368	SAG2132	247	conserved hypothetical protein
ORF02369	SAG2133	118	conserved hypothetical protein
ORF02370	SAG2134	772	membrane protein, putative
ORF02371	SAG2135	179	transcriptional regulator, TetR family, putative
ORF02372	SAG2136	98	conserved hypothetical protein
ORF02373	SAG2137	203	ribosomal protein S4
ORF02374	SAG2138	95	conserved hypothetical protein
ORF02375	SAG2139	451	replicative DNA helicase
ORF02376	SAG2140	150	ribosomal protein L9
ORF02377	SAG2141	660	DHH family protein
ORF02378	SAG2142	613	glucose inhibited division protein A
ORF02379	SAG2143	203	conserved hypothetical protein TIGR00427
ORF02380	SAG2144	373	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
ORF02381	SAG2145	222	L-serine dehydratase, iron-sulfur-dependent, beta subunit
ORF02382	SAG2146	290	L-serine dehydratase, iron-sulfur-dependent, alpha subunit
ORF02383	SAG2147	234	conserved hypothetical protein
ORF02384	SAG2148	179	LysM domain protein
ORF02385	SAG2149	264	cobalt transport family protein
ORF02386	SAG2150	280	ABC transporter, ATP-binding protein
ORF02387	SAG2151	279	ABC transporter, ATP-binding protein, FRAMESHIFT
ORF02388	SAG2152	180	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
ORF02389	SAG2153	427	peptidase, M16 family
ORF02390	SAG2154	414	conserved hypothetical protein
ORF02391	SAG2155	117	conserved hypothetical protein
ORF02392	SAG2156	369	recF protein
ORF02393	SAG2157	278	transporter, putative
ORF02395	SAG2158	220	transcriptional regulator, Cro/Ci family

Table 32: Conversion of ORF Ref Nos. with SAG Ref Nos.

ORF Ref No.	SAGxxxx Ref No.	aa	Annotation
ORF02396	SAG2159	493	Inosine-5'-monophosphate dehydrogenase
ORF02397	SAG2160	161	transcriptional regulator, ArgR family
ORF02398	SAG2161	226	transcriptional regulator, Crp/Fnr family
ORF02399	SAG2162	234	conserved hypothetical protein
ORF02400	SAG2163	410	arginine deiminase
ORF02401	SAG2164	136	acetyltransferase, GNAT family
ORF02402	SAG2165	337	ornithine carbamoyltransferase
ORF02403	SAG2166	475	arginine/ornithine antiporter
ORF02404	SAG2167	318	carbamate kinase
ORF02405	SAG2168	341	tryptophanyl-tRNA synthetase
ORF02406	SAG2169	230	conserved hypothetical protein
ORF02407	SAG2170	290	conserved hypothetical protein
ORF02408	SAG2171	539	ABC transporter, ATP-binding protein
ORF02409	SAG2172	859	ABC transporter, permease protein, putative
ORF02410	SAG2173	159	conserved hypothetical protein TIGR00246
ORF02411	SAG2174	409	serine protease
ORF02412	SAG2175	257	partitioning protein, ParB family
ORF02413	SAG0001	453	chromosomal replication initiator protein DnaA
ORF02415	SAG0002	378	DNA polymerase III, beta subunit
ORF02416	SAG0003	293	diacylglycerol kinase catalytic domain protein, putative
ORF02417	SAG0004	65	conserved hypothetical protein
ORF02418	SAG0005	67	hypothetical protein
ORF02419	SAG0006	371	conserved hypothetical GTP-binding protein
ORF02420	SAG0007	191	peptidyl-tRNA hydrolase
ORF02421	SAG0008	1165	transcription-repair coupling factor
ORF02422	SAG0009	31	hypothetical protein
ORF02423	SAG0010	90	S4 domain protein
ORF02424	SAG0011	123	cell division protein DivIC, putative
ORF02425	SAG0012	44	conserved hypothetical protein
ORF02426	SAG0013	428	conserved hypothetical protein
ORF02427	SAG0014	424	MesJ/Ycf62 family protein
ORF02428	SAG0015	180	hypoxanthine-guanine phosphoribosyltransferase
ORF02429	SAG0016	658	cell division protein FtsH
ORF03000	SAG0157		Dnase-related protein, DEGENERATE
ORF03001	SAG0579	142	conserved hypothetical protein
ORF03002	SAG0580	111	conserved hypothetical protein, truncation
ORF03003	SAG0652		Tn5252, Orf 28 protein, degenerate
ORF03004	SAG0655	57	conserved hypothetical protein
ORF03005	SAG0662	101	cylX protein
ORF03006	SAG0917	83	Tn916, hypothetical protein
ORF03007	SAG0920	23	Tn916, hypothetical protein
ORF03008	SAG0922	61	Tn916, hypothetical protein
ORF03009	SAG0924	28	Tn916, tetM leader peptide
ORF03010	SAG0936	39	Tn916, hypothetical protein
ORF03011	SAG1484	48	ribosomal protein L33
ORF03012	SAG1857	119	HNH endonuclease family protein
ORF03013	SAG1883	128	conserved hypothetical protein
ORF03014	SAG2065	50	ribosomal protein L33
ORF03015	SAG2004	67	conjugal transfer protein, Interruption-N

Tabl 33: List of GAS ORFs which are shared with GBS and Spn

gi 13621326 gb AAK33146.1	gi 13621393 gb AAK33207.1
gi 13621327 gb AAK33147.1	gi 13621394 gb AAK33208.1
gi 13621328 gb AAK33148.1	gi 13621397 gb AAK33210.1
gi 13621329 gb AAK33149.1	gi 13621398 gb AAK33211.1
gi 13621330 gb AAK33150.1	gi 13621399 gb AAK33212.1
gi 13621331 gb AAK33151.1	gi 13621401 gb AAK33214.1
gi 13621332 gb AAK33152.1	gi 13621403 gb AAK33215.1
gi 13621333 gb AAK33153.1	gi 13621404 gb AAK33216.1
gi 13621334 gb AAK33154.1	gi 13621405 gb AAK33217.1
gi 13621335 gb AAK33155.1	gi 13621407 gb AAK33218.1
gi 13621337 gb AAK33156.1	gi 13621408 gb AAK33219.1
gi 13621340 gb AAK33158.1	gi 13621409 gb AAK33220.1
gi 13621341 gb AAK33159.1	gi 13621413 gb AAK33224.1
gi 13621343 gb AAK33160.1	gi 13621415 gb AAK33226.1
gi 13621344 gb AAK33161.1	gi 13621416 gb AAK33227.1
gi 13621346 gb AAK33163.1	gi 13621418 gb AAK33229.1
gi 13621347 gb AAK33164.1	gi 13621419 gb AAK33230.1
gi 13621348 gb AAK33165.1	gi 13621424 gb AAK33234.1
gi 13621349 gb AAK33166.1	gi 13621425 gb AAK33235.1
gi 13621350 gb AAK33167.1	gi 13621426 gb AAK33236.1
gi 13621353 gb AAK33169.1	gi 13621434 gb AAK33243.1
gi 13621354 gb AAK33170.1	gi 13621450 gb AAK33258.1
gi 13621355 gb AAK33171.1	gi 13621455 gb AAK33262.1
gi 13621357 gb AAK33173.1	gi 13621456 gb AAK33263.1
gi 13621358 gb AAK33174.1	gi 13621457 gb AAK33264.1
gi 13621359 gb AAK33175.1	gi 13621467 gb AAK33273.1
gi 13621361 gb AAK33176.1	gi 13621468 gb AAK33274.1
gi 13621362 gb AAK33177.1	gi 13621469 gb AAK33275.1
gi 13621363 gb AAK33178.1	gi 13621470 gb AAK33276.1
gi 13621364 gb AAK33179.1	gi 13621471 gb AAK33277.1
gi 13621365 gb AAK33180.1	gi 13621472 gb AAK33278.1
gi 13621366 gb AAK33181.1	gi 13621473 gb AAK33279.1
gi 13621367 gb AAK33182.1	gi 13621476 gb AAK33281.1
gi 13621368 gb AAK33183.1	gi 13621477 gb AAK33282.1
gi 13621369 gb AAK33184.1	gi 13621478 gb AAK33283.1
gi 13621370 gb AAK33185.1	gi 13621480 gb AAK33285.1
gi 13621372 gb AAK33186.1	gi 13621481 gb AAK33286.1
gi 13621373 gb AAK33187.1	gi 13621491 gb AAK33295.1
gi 13621374 gb AAK33188.1	gi 13621494 gb AAK33298.1
gi 13621375 gb AAK33189.1	gi 13621496 gb AAK33299.1
gi 13621376 gb AAK33190.1	gi 13621501 gb AAK33304.1
gi 13621377 gb AAK33191.1	gi 13621502 gb AAK33305.1
gi 13621378 gb AAK33192.1	gi 13621505 gb AAK33307.1
gi 13621379 gb AAK33193.1	gi 13621506 gb AAK33308.1
gi 13621380 gb AAK33194.1	gi 13621507 gb AAK33309.1
gi 13621382 gb AAK33196.1	gi 13621510 gb AAK33312.1
gi 13621383 gb AAK33197.1	gi 13621511 gb AAK33313.1
gi 13621384 gb AAK33198.1	gi 13621513 gb AAK33315.1
gi 13621385 gb AAK33199.1	gi 13621516 gb AAK33317.1
gi 13621386 gb AAK33200.1	gi 13621518 gb AAK33319.1
gi 13621387 gb AAK33201.1	gi 13621521 gb AAK33322.1
gi 13621388 gb AAK33202.1	gi 13621522 gb AAK33323.1
gi 13621389 gb AAK33203.1	gi 13621523 gb AAK33324.1
gi 13621390 gb AAK33204.1	gi 13621524 gb AAK33325.1
gi 13621391 gb AAK33205.1	gi 13621525 gb AAK33326.1
gi 13621392 gb AAK33206.1	gi 13621527 gb AAK33327.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13621528 gb AAK33328.1	gi 13621595 gb AAK33389.1
gi 13621529 gb AAK33329.1	gi 13621596 gb AAK33390.1
gi 13621530 gb AAK33330.1	gi 13621597 gb AAK33391.1
gi 13621531 gb AAK33331.1	gi 13621598 gb AAK33392.1
gi 13621532 gb AAK33332.1	gi 13621599 gb AAK33393.1
gi 13621533 gb AAK33333.1	gi 13621600 gb AAK33394.1
gi 13621534 gb AAK33334.1	gi 13621602 gb AAK33395.1
gi 13621535 gb AAK33335.1	gi 13621603 gb AAK33396.1
gi 13621536 gb AAK33336.1	gi 13621604 gb AAK33397.1
gi 13621537 gb AAK33337.1	gi 13621605 gb AAK33398.1
gi 13621539 gb AAK33338.1	gi 13621606 gb AAK33399.1
gi 13621540 gb AAK33339.1	gi 13621607 gb AAK33400.1
gi 13621541 gb AAK33340.1	gi 13621608 gb AAK33401.1
gi 13621542 gb AAK33341.1	gi 13621609 gb AAK33402.1
gi 13621543 gb AAK33342.1	gi 13621611 gb AAK33404.1
gi 13621544 gb AAK33343.1	gi 13621614 gb AAK33406.1
gi 13621546 gb AAK33345.1	gi 13621615 gb AAK33407.1
gi 13621547 gb AAK33346.1	gi 13621616 gb AAK33408.1
gi 13621548 gb AAK33347.1	gi 13621617 gb AAK33409.1
gi 13621550 gb AAK33348.1	gi 13621618 gb AAK33410.1
gi 13621551 gb AAK33349.1	gi 13621619 gb AAK33411.1
gi 13621552 gb AAK33350.1	gi 13621620 gb AAK33412.1
gi 13621553 gb AAK33351.1	gi 13621621 gb AAK33413.1
gi 13621554 gb AAK33352.1	gi 13621622 gb AAK33414.1
gi 13621555 gb AAK33353.1	gi 13621623 gb AAK33415.1
gi 13621557 gb AAK33355.1	gi 13621624 gb AAK33416.1
gi 13621559 gb AAK33356.1	gi 13621625 gb AAK33417.1
gi 13621560 gb AAK33357.1	gi 13621627 gb AAK33419.1
gi 13621561 gb AAK33358.1	gi 13621629 gb AAK33420.1
gi 13621562 gb AAK33359.1	gi 13621630 gb AAK33421.1
gi 13621563 gb AAK33360.1	gi 13621631 gb AAK33422.1
gi 13621564 gb AAK33361.1	gi 13621633 gb AAK33424.1
gi 13621565 gb AAK33362.1	gi 13621634 gb AAK33425.1
gi 13621566 gb AAK33363.1	gi 13621636 gb AAK33427.1
gi 13621567 gb AAK33364.1	gi 13621637 gb AAK33428.1
gi 13621569 gb AAK33365.1	gi 13621638 gb AAK33429.1
gi 13621571 gb AAK33367.1	gi 13621640 gb AAK33430.1
gi 13621572 gb AAK33368.1	gi 13621642 gb AAK33432.1
gi 13621573 gb AAK33369.1	gi 13621644 gb AAK33434.1
gi 13621574 gb AAK33370.1	gi 13621645 gb AAK33435.1
gi 13621575 gb AAK33371.1	gi 13621647 gb AAK33437.1
gi 13621576 gb AAK33372.1	gi 13621648 gb AAK33438.1
gi 13621577 gb AAK33373.1	gi 13621650 gb AAK33440.1
gi 13621579 gb AAK33374.1	gi 13621651 gb AAK33441.1
gi 13621581 gb AAK33376.1	gi 13621652 gb AAK33442.1
gi 13621582 gb AAK33377.1	gi 13621657 gb AAK33446.1
gi 13621583 gb AAK33378.1	gi 13621658 gb AAK33447.1
gi 13621584 gb AAK33379.1	gi 13621660 gb AAK33449.1
gi 13621585 gb AAK33380.1	gi 13621670 gb AAK33458.1
gi 13621586 gb AAK33381.1	gi 13621671 gb AAK33459.1
gi 13621588 gb AAK33383.1	gi 13621672 gb AAK33460.1
gi 13621589 gb AAK33384.1	gi 13621675 gb AAK33462.1
gi 13621590 gb AAK33385.1	gi 13621676 gb AAK33463.1
gi 13621592 gb AAK33386.1	gi 13621678 gb AAK33465.1
gi 13621593 gb AAK33387.1	gi 13621680 gb AAK33467.1
gi 13621594 gb AAK33388.1	gi 13621681 gb AAK33468.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13621682 gb AAK33469.1	gi 13621796 gb AAK33573.1
gi 13621683 gb AAK33470.1	gi 13621797 gb AAK33574.1
gi 13621684 gb AAK33471.1	gi 13621799 gb AAK33576.1
gi 13621685 gb AAK33472.1	gi 13621800 gb AAK33577.1
gi 13621688 gb AAK33474.1	gi 13621802 gb AAK33579.1
gi 13621689 gb AAK33475.1	gi 13621806 gb AAK33583.1
gi 13621690 gb AAK33476.1	gi 13621808 gb AAK33584.1
gi 13621691 gb AAK33477.1	gi 13621809 gb AAK33585.1
gi 13621692 gb AAK33478.1	gi 13621810 gb AAK33586.1
gi 13621693 gb AAK33479.1	gi 13621811 gb AAK33587.1
gi 13621694 gb AAK33480.1	gi 13621812 gb AAK33588.1
gi 13621695 gb AAK33481.1	gi 13621813 gb AAK33589.1
gi 13621697 gb AAK33483.1	gi 13621814 gb AAK33590.1
gi 13621698 gb AAK33484.1	gi 13621817 gb AAK33592.1
gi 13621700 gb AAK33485.1	gi 13621818 gb AAK33593.1
gi 13621701 gb AAK33486.1	gi 13621819 gb AAK33594.1
gi 13621702 gb AAK33487.1	gi 13621820 gb AAK33595.1
gi 13621714 gb AAK33498.1	gi 13621821 gb AAK33596.1
gi 13621715 gb AAK33499.1	gi 13621822 gb AAK33597.1
gi 13621717 gb AAK33501.1	gi 13621823 gb AAK33598.1
gi 13621718 gb AAK33502.1	gi 13621824 gb AAK33599.1
gi 13621719 gb AAK33503.1	gi 13621825 gb AAK33600.1
gi 13621720 gb AAK33504.1	gi 13621826 gb AAK33601.1
gi 13621726 gb AAK33509.1	gi 13621828 gb AAK33602.1
gi 13621727 gb AAK33510.1	gi 13621829 gb AAK33603.1
gi 13621729 gb AAK33512.1	gi 13621830 gb AAK33604.1
gi 13621730 gb AAK33513.1	gi 13621831 gb AAK33605.1
gi 13621731 gb AAK33514.1	gi 13621834 gb AAK33608.1
gi 13621732 gb AAK33515.1	gi 13621835 gb AAK33609.1
gi 13621733 gb AAK33516.1	gi 13621836 gb AAK33610.1
gi 13621734 gb AAK33517.1	gi 13621837 gb AAK33611.1
gi 13621735 gb AAK33518.1	gi 13621839 gb AAK33612.1
gi 13621736 gb AAK33519.1	gi 13621840 gb AAK33613.1
gi 13621741 gb AAK33523.1	gi 13621841 gb AAK33614.1
gi 13621742 gb AAK33524.1	gi 13621842 gb AAK33615.1
gi 13621743 gb AAK33525.1	gi 13621843 gb AAK33616.1
gi 13621744 gb AAK33526.1	gi 13621844 gb AAK33617.1
gi 13621745 gb AAK33527.1	gi 13621898 gb AAK33667.1
gi 13621747 gb AAK33528.1	gi 13621901 gb AAK33670.1
gi 13621756 gb AAK33537.1	gi 13621902 gb AAK33671.1
gi 13621773 gb AAK33552.1	gi 13621904 gb AAK33672.1
gi 13621774 gb AAK33553.1	gi 13621907 gb AAK33675.1
gi 13621775 gb AAK33554.1	gi 13621908 gb AAK33676.1
gi 13621777 gb AAK33556.1	gi 13621909 gb AAK33677.1
gi 13621778 gb AAK33557.1	gi 13621910 gb AAK33678.1
gi 13621779 gb AAK33558.1	gi 13621912 gb AAK33680.1
gi 13621781 gb AAK33559.1	gi 13621924 gb AAK33690.1
gi 13621782 gb AAK33560.1	gi 13621929 gb AAK33694.1
gi 13621785 gb AAK33563.1	gi 13621930 gb AAK33695.1
gi 13621786 gb AAK33564.1	gi 13621931 gb AAK33696.1
gi 13621787 gb AAK33565.1	gi 13621933 gb AAK33698.1
gi 13621788 gb AAK33566.1	gi 13621934 gb AAK33699.1
gi 13621789 gb AAK33567.1	gi 13621935 gb AAK33700.1
gi 13621790 gb AAK33568.1	gi 13621936 gb AAK33701.1
gi 13621793 gb AAK33571.1	gi 13621937 gb AAK33702.1
gi 13621794 gb AAK33572.1	gi 13621938 gb AAK33703.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13621939 gb AAK33704.1	gi 13622034 gb AAK33790.1
gi 13621942 gb AAK33706.1	gi 13622035 gb AAK33791.1
gi 13621944 gb AAK33708.1	gi 13622039 gb AAK33794.1
gi 13621945 gb AAK33709.1	gi 13622041 gb AAK33796.1
gi 13621946 gb AAK33710.1	gi 13622042 gb AAK33797.1
gi 13621950 gb AAK33714.1	gi 13622043 gb AAK33798.1
gi 13621953 gb AAK33716.1	gi 13622044 gb AAK33799.1
gi 13621954 gb AAK33717.1	gi 13622045 gb AAK33800.1
gi 13621955 gb AAK33718.1	gi 13622046 gb AAK33801.1
gi 13621956 gb AAK33719.1	gi 13622048 gb AAK33802.1
gi 13621957 gb AAK33720.1	gi 13622049 gb AAK33803.1
gi 13621958 gb AAK33721.1	gi 13622050 gb AAK33804.1
gi 13621959 gb AAK33722.1	gi 13622051 gb AAK33805.1
gi 13621961 gb AAK33723.1	gi 13622052 gb AAK33806.1
gi 13621975 gb AAK33736.1	gi 13622054 gb AAK33808.1
gi 13621977 gb AAK33738.1	gi 13622055 gb AAK33809.1
gi 13621978 gb AAK33739.1	gi 13622056 gb AAK33810.1
gi 13621979 gb AAK33740.1	gi 13622058 gb AAK33812.1
gi 13621980 gb AAK33741.1	gi 13622060 gb AAK33813.1
gi 13621981 gb AAK33742.1	gi 13622062 gb AAK33815.1
gi 13621982 gb AAK33743.1	gi 13622064 gb AAK33817.1
gi 13621985 gb AAK33745.1	gi 13622065 gb AAK33818.1
gi 13621986 gb AAK33746.1	gi 13622068 gb AAK33821.1
gi 13621987 gb AAK33747.1	gi 13622069 gb AAK33822.1
gi 13621989 gb AAK33749.1	gi 13622070 gb AAK33823.1
gi 13621990 gb AAK33750.1	gi 13622071 gb AAK33824.1
gi 13621992 gb AAK33752.1	gi 13622073 gb AAK33825.1
gi 13621993 gb AAK33753.1	gi 13622074 gb AAK33826.1
gi 13621994 gb AAK33754.1	gi 13622075 gb AAK33827.1
gi 13621996 gb AAK33755.1	gi 13622077 gb AAK33829.1
gi 13621997 gb AAK33756.1	gi 13622079 gb AAK33831.1
gi 13621998 gb AAK33757.1	gi 13622083 gb AAK33834.1
gi 13621999 gb AAK33758.1	gi 13622085 gb AAK33836.1
gi 13622000 gb AAK33759.1	gi 13622086 gb AAK33837.1
gi 13622001 gb AAK33760.1	gi 13622087 gb AAK33838.1
gi 13622002 gb AAK33761.1	gi 13622088 gb AAK33839.1
gi 13622003 gb AAK33762.1	gi 13622089 gb AAK33840.1
gi 13622004 gb AAK33763.1	gi 13622090 gb AAK33841.1
gi 13622005 gb AAK33764.1	gi 13622091 gb AAK33842.1
gi 13622006 gb AAK33765.1	gi 13622092 gb AAK33843.1
gi 13622008 gb AAK33766.1	gi 13622093 gb AAK33844.1
gi 13622009 gb AAK33767.1	gi 13622095 gb AAK33845.1
gi 13622010 gb AAK33768.1	gi 13622096 gb AAK33846.1
gi 13622012 gb AAK33770.1	gi 13622097 gb AAK33847.1
gi 13622013 gb AAK33771.1	gi 13622162 gb AAK33908.1
gi 13622017 gb AAK33774.1	gi 13622163 gb AAK33909.1
gi 13622018 gb AAK33775.1	gi 13622164 gb AAK33910.1
gi 13622019 gb AAK33776.1	gi 13622165 gb AAK33911.1
gi 13622020 gb AAK33777.1	gi 13622166 gb AAK33912.1
gi 13622021 gb AAK33778.1	gi 13622169 gb AAK33914.1
gi 13622024 gb AAK33781.1	gi 13622170 gb AAK33915.1
gi 13622025 gb AAK33782.1	gi 13622171 gb AAK33916.1
gi 13622026 gb AAK33783.1	gi 13622172 gb AAK33917.1
gi 13622031 gb AAK33787.1	gi 13622174 gb AAK33919.1
gi 13622032 gb AAK33788.1	gi 13622175 gb AAK33920.1
gi 13622033 gb AAK33789.1	gi 13622176 gb AAK33921.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13622177 gb AAK33922.1	gi 13622269 gb AAK34006.1
gi 13622179 gb AAK33923.1	gi 13622271 gb AAK34007.1
gi 13622180 gb AAK33924.1	gi 13622272 gb AAK34008.1
gi 13622181 gb AAK33925.1	gi 13622273 gb AAK34009.1
gi 13622182 gb AAK33926.1	gi 13622274 gb AAK34010.1
gi 13622183 gb AAK33927.1	gi 13622275 gb AAK34011.1
gi 13622184 gb AAK33928.1	gi 13622276 gb AAK34012.1
gi 13622185 gb AAK33929.1	gi 13622277 gb AAK34013.1
gi 13622186 gb AAK33930.1	gi 13622278 gb AAK34014.1
gi 13622189 gb AAK33932.1	gi 13622279 gb AAK34015.1
gi 13622190 gb AAK33933.1	gi 13622281 gb AAK34017.1
gi 13622191 gb AAK33934.1	gi 13622282 gb AAK34018.1
gi 13622192 gb AAK33935.1	gi 13622283 gb AAK34019.1
gi 13622198 gb AAK33940.1	gi 13622284 gb AAK34020.1
gi 13622200 gb AAK33942.1	gi 13622285 gb AAK34021.1
gi 13622201 gb AAK33943.1	gi 13622287 gb AAK34022.1
gi 13622204 gb AAK33946.1	gi 13622288 gb AAK34023.1
gi 13622205 gb AAK33947.1	gi 13622289 gb AAK34024.1
gi 13622207 gb AAK33949.1	gi 13622290 gb AAK34025.1
gi 13622208 gb AAK33950.1	gi 13622294 gb AAK34029.1
gi 13622211 gb AAK33952.1	gi 13622295 gb AAK34030.1
gi 13622213 gb AAK33954.1	gi 13622296 gb AAK34031.1
gi 13622214 gb AAK33955.1	gi 13622297 gb AAK34032.1
gi 13622215 gb AAK33956.1	gi 13622298 gb AAK34033.1
gi 13622216 gb AAK33957.1	gi 13622299 gb AAK34034.1
gi 13622217 gb AAK33958.1	gi 13622301 gb AAK34035.1
gi 13622218 gb AAK33959.1	gi 13622306 gb AAK34040.1
gi 13622219 gb AAK33960.1	gi 13622326 gb AAK34058.1
gi 13622222 gb AAK33962.1	gi 13622328 gb AAK34060.1
gi 13622223 gb AAK33963.1	gi 13622329 gb AAK34061.1
gi 13622224 gb AAK33964.1	gi 13622330 gb AAK34062.1
gi 13622233 gb AAK33972.1	gi 13622332 gb AAK34064.1
gi 13622235 gb AAK33974.1	gi 13622333 gb AAK34065.1
gi 13622236 gb AAK33975.1	gi 13622335 gb AAK34066.1
gi 13622237 gb AAK33976.1	gi 13622338 gb AAK34069.1
gi 13622239 gb AAK33978.1	gi 13622339 gb AAK34070.1
gi 13622240 gb AAK33979.1	gi 13622340 gb AAK34071.1
gi 13622241 gb AAK33980.1	gi 13622341 gb AAK34072.1
gi 13622242 gb AAK33981.1	gi 13622343 gb AAK34073.1
gi 13622243 gb AAK33982.1	gi 13622350 gb AAK34080.1
gi 13622244 gb AAK33983.1	gi 13622351 gb AAK34081.1
gi 13622250 gb AAK33988.1	gi 13622352 gb AAK34082.1
gi 13622252 gb AAK33990.1	gi 13622353 gb AAK34083.1
gi 13622253 gb AAK33991.1	gi 13622355 gb AAK34084.1
gi 13622255 gb AAK33993.1	gi 13622356 gb AAK34085.1
gi 13622256 gb AAK33994.1	gi 13622357 gb AAK34086.1
gi 13622257 gb AAK33995.1	gi 13622358 gb AAK34087.1
gi 13622259 gb AAK33996.1	gi 13622359 gb AAK34088.1
gi 13622260 gb AAK33997.1	gi 13622360 gb AAK34089.1
gi 13622261 gb AAK33998.1	gi 13622361 gb AAK34090.1
gi 13622262 gb AAK33999.1	gi 13622362 gb AAK34091.1
gi 13622263 gb AAK34000.1	gi 13622363 gb AAK34092.1
gi 13622264 gb AAK34001.1	gi 13622364 gb AAK34093.1
gi 13622265 gb AAK34002.1	gi 13622366 gb AAK34094.1
gi 13622266 gb AAK34003.1	gi 13622367 gb AAK34095.1
gi 13622268 gb AAK34005.1	gi 13622368 gb AAK34096.1

Tabl 33: List f GAS ORFs which are shar d with GBS and Spn

gi 13622369 gb AAK34097.1	gi 13622471 gb AAK34189.1
gi 13622370 gb AAK34098.1	gi 13622473 gb AAK34191.1
gi 13622371 gb AAK34099.1	gi 13622474 gb AAK34192.1
gi 13622372 gb AAK34100.1	gi 13622477 gb AAK34195.1
gi 13622373 gb AAK34101.1	gi 13622478 gb AAK34196.1
gi 13622374 gb AAK34102.1	gi 13622479 gb AAK34197.1
gi 13622375 gb AAK34103.1	gi 13622481 gb AAK34198.1
gi 13622376 gb AAK34104.1	gi 13622482 gb AAK34199.1
gi 13622377 gb AAK34105.1	gi 13622483 gb AAK34200.1
gi 13622378 gb AAK34106.1	gi 13622484 gb AAK34201.1
gi 13622380 gb AAK34107.1	gi 13622485 gb AAK34202.1
gi 13622383 gb AAK34110.1	gi 13622486 gb AAK34203.1
gi 13622384 gb AAK34111.1	gi 13622491 gb AAK34207.1
gi 13622387 gb AAK34114.1	gi 13622492 gb AAK34208.1
gi 13622389 gb AAK34116.1	gi 13622493 gb AAK34209.1
gi 13622394 gb AAK34120.1	gi 13622494 gb AAK34210.1
gi 13622395 gb AAK34121.1	gi 13622495 gb AAK34211.1
gi 13622396 gb AAK34122.1	gi 13622496 gb AAK34212.1
gi 13622398 gb AAK34124.1	gi 13622497 gb AAK34213.1
gi 13622399 gb AAK34125.1	gi 13622499 gb AAK34214.1
gi 13622400 gb AAK34126.1	gi 13622500 gb AAK34215.1
gi 13622401 gb AAK34127.1	gi 13622501 gb AAK34216.1
gi 13622403 gb AAK34128.1	gi 13622506 gb AAK34221.1
gi 13622405 gb AAK34130.1	gi 13622507 gb AAK34222.1
gi 13622406 gb AAK34131.1	gi 13622508 gb AAK34223.1
gi 13622407 gb AAK34132.1	gi 13622509 gb AAK34224.1
gi 13622408 gb AAK34133.1	gi 13622511 gb AAK34225.1
gi 13622415 gb AAK34139.1	gi 13622512 gb AAK34226.1
gi 13622416 gb AAK34140.1	gi 13622513 gb AAK34227.1
gi 13622417 gb AAK34141.1	gi 13622515 gb AAK34229.1
gi 13622419 gb AAK34143.1	gi 13622516 gb AAK34230.1
gi 13622420 gb AAK34144.1	gi 13622517 gb AAK34231.1
gi 13622424 gb AAK34147.1	gi 13622518 gb AAK34232.1
gi 13622425 gb AAK34148.1	gi 13622520 gb AAK34233.1
gi 13622431 gb AAK34153.1	gi 13622521 gb AAK34234.1
gi 13622432 gb AAK34154.1	gi 13622523 gb AAK34236.1
gi 13622433 gb AAK34155.1	gi 13622524 gb AAK34237.1
gi 13622434 gb AAK34156.1	gi 13622525 gb AAK34238.1
gi 13622435 gb AAK34157.1	gi 13622526 gb AAK34239.1
gi 13622436 gb AAK34158.1	gi 13622527 gb AAK34240.1
gi 13622437 gb AAK34159.1	gi 13622579 gb AAK34289.1
gi 13622444 gb AAK34165.1	gi 13622583 gb AAK34292.1
gi 13622447 gb AAK34168.1	gi 13622585 gb AAK34294.1
gi 13622450 gb AAK34170.1	gi 13622587 gb AAK34296.1
gi 13622451 gb AAK34171.1	gi 13622588 gb AAK34297.1
gi 13622455 gb AAK34175.1	gi 13622590 gb AAK34299.1
gi 13622457 gb AAK34177.1	gi 13622591 gb AAK34300.1
gi 13622458 gb AAK34178.1	gi 13622593 gb AAK34301.1
gi 13622460 gb AAK34179.1	gi 13622595 gb AAK34303.1
gi 13622461 gb AAK34180.1	gi 13622596 gb AAK34304.1
gi 13622462 gb AAK34181.1	gi 13622597 gb AAK34305.1
gi 13622463 gb AAK34182.1	gi 13622598 gb AAK34306.1
gi 13622464 gb AAK34183.1	gi 13622599 gb AAK34307.1
gi 13622465 gb AAK34184.1	gi 13622600 gb AAK34308.1
gi 13622467 gb AAK34186.1	gi 13622601 gb AAK34309.1
gi 13622468 gb AAK34187.1	gi 13622603 gb AAK34310.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13622604 gb AAK34311.1	gi 13622711 gb AAK34408.1
gi 13622606 gb AAK34313.1	gi 13622713 gb AAK34410.1
gi 13622607 gb AAK34314.1	gi 13622714 gb AAK34411.1
gi 13622608 gb AAK34315.1	gi 13622715 gb AAK34412.1
gi 13622609 gb AAK34316.1	gi 13622718 gb AAK34414.1
gi 13622610 gb AAK34317.1	gi 13622719 gb AAK34415.1
gi 13622611 gb AAK34318.1	gi 13622720 gb AAK34416.1
gi 13622612 gb AAK34319.1	gi 13622721 gb AAK34417.1
gi 13622615 gb AAK34321.1	gi 13622722 gb AAK34418.1
gi 13622616 gb AAK34322.1	gi 13622723 gb AAK34419.1
gi 13622617 gb AAK34323.1	gi 13622727 gb AAK34422.1
gi 13622618 gb AAK34324.1	gi 13622728 gb AAK34423.1
gi 13622621 gb AAK34327.1	gi 13622729 gb AAK34424.1
gi 13622622 gb AAK34328.1	gi 13622730 gb AAK34425.1
gi 13622623 gb AAK34329.1	gi 13622731 gb AAK34426.1
gi 13622624 gb AAK34330.1	gi 13622733 gb AAK34428.1
gi 13622625 gb AAK34331.1	gi 13622734 gb AAK34429.1
gi 13622626 gb AAK34332.1	gi 13622735 gb AAK34430.1
gi 13622628 gb AAK34333.1	gi 13622736 gb AAK34431.1
gi 13622629 gb AAK34334.1	gi 13622737 gb AAK34432.1
gi 13622630 gb AAK34335.1	gi 13622740 gb AAK34434.1
gi 13622631 gb AAK34336.1	gi 13622741 gb AAK34435.1
gi 13622632 gb AAK34337.1	gi 13622742 gb AAK34436.1
gi 13622634 gb AAK34339.1	gi 13622744 gb AAK34438.1
gi 13622636 gb AAK34341.1	gi 13622745 gb AAK34439.1
gi 13622640 gb AAK34344.1	gi 13622746 gb AAK34440.1
gi 13622641 gb AAK34345.1	gi 13622749 gb AAK34442.1
gi 13622652 gb AAK34355.1	gi 13622750 gb AAK34443.1
gi 13622653 gb AAK34356.1	gi 13622751 gb AAK34444.1
gi 13622654 gb AAK34357.1	gi 13622752 gb AAK34445.1
gi 13622656 gb AAK34359.1	gi 13622753 gb AAK34446.1
gi 13622660 gb AAK34363.1	gi 13622754 gb AAK34447.1
gi 13622665 gb AAK34367.1	gi 13622760 gb AAK34452.1
gi 13622668 gb AAK34370.1	gi 13622762 gb AAK34454.1
gi 13622675 gb AAK34376.1	gi 13622763 gb AAK34455.1
gi 13622676 gb AAK34377.1	gi 13622764 gb AAK34456.1
gi 13622683 gb AAK34383.1	gi 13622765 gb AAK34457.1
gi 13622684 gb AAK34384.1	gi 13622766 gb AAK34458.1
gi 13622685 gb AAK34385.1	gi 13622767 gb AAK34459.1
gi 13622688 gb AAK34387.1	gi 13622768 gb AAK34460.1
gi 13622689 gb AAK34388.1	gi 13622770 gb AAK34462.1
gi 13622690 gb AAK34389.1	gi 13622771 gb AAK34463.1
gi 13622691 gb AAK34390.1	gi 13622774 gb AAK34465.1
gi 13622692 gb AAK34391.1	gi 13622775 gb AAK34466.1
gi 13622693 gb AAK34392.1	gi 13622776 gb AAK34467.1
gi 13622694 gb AAK34393.1	gi 13622777 gb AAK34468.1
gi 13622695 gb AAK34394.1	gi 13622778 gb AAK34469.1
gi 13622696 gb AAK34395.1	gi 13622779 gb AAK34470.1
gi 13622698 gb AAK34396.1	gi 13622780 gb AAK34471.1
gi 13622699 gb AAK34397.1	gi 13622781 gb AAK34472.1
gi 13622700 gb AAK34398.1	gi 13622782 gb AAK34473.1
gi 13622701 gb AAK34399.1	gi 13622783 gb AAK34474.1
gi 13622702 gb AAK34400.1	gi 13622785 gb AAK34475.1
gi 13622703 gb AAK34401.1	gi 13622787 gb AAK34477.1
gi 13622704 gb AAK34402.1	gi 13622789 gb AAK34479.1
gi 13622705 gb AAK34403.1	gi 13622790 gb AAK34480.1

Table 33: List of GAS ORFs which are shared with GBS and Spn

gi 13622791 gb AAK34481.1	gi 13622870 gb AAK34553.1
gi 13622792 gb AAK34482.1	gi 13622873 gb AAK34555.1
gi 13622793 gb AAK34483.1	gi 13622875 gb AAK34557.1
gi 13622794 gb AAK34484.1	gi 13622876 gb AAK34558.1
gi 13622795 gb AAK34485.1	gi 13622877 gb AAK34559.1
gi 13622796 gb AAK34486.1	gi 13622878 gb AAK34560.1
gi 13622798 gb AAK34487.1	gi 13622879 gb AAK34561.1
gi 13622799 gb AAK34488.1	gi 13622880 gb AAK34562.1
gi 13622800 gb AAK34489.1	gi 13622881 gb AAK34563.1
gi 13622801 gb AAK34490.1	gi 13622882 gb AAK34564.1
gi 13622802 gb AAK34491.1	gi 13622885 gb AAK34566.1
gi 13622803 gb AAK34492.1	gi 13622886 gb AAK34567.1
gi 13622804 gb AAK34493.1	gi 13622887 gb AAK34568.1
gi 13622805 gb AAK34494.1	gi 13622888 gb AAK34569.1
gi 13622806 gb AAK34495.1	gi 13622890 gb AAK34571.1
gi 13622807 gb AAK34496.1	gi 13622893 gb AAK34574.1
gi 13622808 gb AAK34497.1	gi 13622896 gb AAK34576.1
gi 13622809 gb AAK34498.1	gi 13622898 gb AAK34578.1
gi 13622810 gb AAK34499.1	gi 13622899 gb AAK34579.1
gi 13622812 gb AAK34500.1	gi 13622900 gb AAK34580.1
gi 13622813 gb AAK34501.1	gi 13622901 gb AAK34581.1
gi 13622814 gb AAK34502.1	gi 13622903 gb AAK34583.1
gi 13622815 gb AAK34503.1	gi 13622905 gb AAK34585.1
gi 13622818 gb AAK34506.1	gi 13622906 gb AAK34586.1
gi 13622821 gb AAK34509.1	gi 13622907 gb AAK34587.1
gi 13622822 gb AAK34510.1	gi 13622908 gb AAK34588.1
gi 13622823 gb AAK34511.1	gi 13622910 gb AAK34589.1
gi 13622825 gb AAK34512.1	gi 13622911 gb AAK34590.1
gi 13622826 gb AAK34513.1	gi 13622912 gb AAK34591.1
gi 13622827 gb AAK34514.1	gi 13622913 gb AAK34592.1
gi 13622828 gb AAK34515.1	gi 13622914 gb AAK34593.1
gi 13622829 gb AAK34516.1	gi 13622915 gb AAK34594.1
gi 13622830 gb AAK34517.1	gi 13622917 gb AAK34596.1
gi 13622833 gb AAK34520.1	gi 13622918 gb AAK34597.1
gi 13622838 gb AAK34524.1	gi 13622919 gb AAK34598.1
gi 13622839 gb AAK34525.1	gi 13622921 gb AAK34599.1
gi 13622840 gb AAK34526.1	gi 13622922 gb AAK34600.1
gi 13622841 gb AAK34527.1	gi 13622924 gb AAK34602.1
gi 13622847 gb AAK34532.1	gi 13622925 gb AAK34603.1
gi 13622848 gb AAK34533.1	gi 13622926 gb AAK34604.1
gi 13622849 gb AAK34534.1	gi 13622927 gb AAK34605.1
gi 13622853 gb AAK34537.1	gi 13622928 gb AAK34606.1
gi 13622854 gb AAK34538.1	gi 13622929 gb AAK34607.1
gi 13622856 gb AAK34540.1	gi 13622930 gb AAK34608.1
gi 13622857 gb AAK34541.1	gi 13622931 gb AAK34609.1
gi 13622858 gb AAK34542.1	gi 13622933 gb AAK34610.1
gi 13622860 gb AAK34543.1	gi 13622941 gb AAK34617.1
gi 13622861 gb AAK34544.1	gi 13622944 gb AAK34620.1
gi 13622862 gb AAK34545.1	gi 13622945 gb AAK34621.1
gi 13622863 gb AAK34546.1	gi 13622947 gb AAK34623.1
gi 13622864 gb AAK34547.1	gi 13622948 gb AAK34624.1
gi 13622865 gb AAK34548.1	gi 13622949 gb AAK34625.1
gi 13622866 gb AAK34549.1	gi 13622950 gb AAK34626.1
gi 13622867 gb AAK34550.1	gi 13622952 gb AAK34627.1
gi 13622868 gb AAK34551.1	gi 13622955 gb AAK34630.1
gi 13622869 gb AAK34552.1	gi 13622956 gb AAK34631.1

Tabl 34: List of GAS ORF's which are shared with GBS but not with Spn

gi 13621381 gb AAK33195.1	gi 13621988 gb AAK33748.1
gi 13621423 gb AAK33233.1	gi 13622014 gb AAK33772.1
gi 13621440 gb AAK33249.1	gi 13622015 gb AAK33773.1
gi 13621443 gb AAK33251.1	gi 13622022 gb AAK33779.1
gi 13621453 gb AAK33260.1	gi 13622023 gb AAK33780.1
gi 13621454 gb AAK33261.1	gi 13622028 gb AAK33784.1
gi 13621479 gb AAK33284.1	gi 13622029 gb AAK33785.1
gi 13621482 gb AAK33287.1	gi 13622037 gb AAK33792.1
gi 13621492 gb AAK33296.1	gi 13622038 gb AAK33793.1
gi 13621493 gb AAK33297.1	gi 13622040 gb AAK33795.1
gi 13621497 gb AAK33300.1	gi 13622057 gb AAK33811.1
gi 13621498 gb AAK33301.1	gi 13622061 gb AAK33814.1
gi 13621512 gb AAK33314.1	gi 13622063 gb AAK33816.1
gi 13621514 gb AAK33316.1	gi 13622066 gb AAK33819.1
gi 13621556 gb AAK33354.1	gi 13622067 gb AAK33820.1
gi 13621570 gb AAK33366.1	gi 13622076 gb AAK33828.1
gi 13621587 gb AAK33382.1	gi 13622078 gb AAK33830.1
gi 13621610 gb AAK33403.1	gi 13622084 gb AAK33835.1
gi 13621613 gb AAK33405.1	gi 13622098 gb AAK33848.1
gi 13621626 gb AAK33418.1	gi 13622099 gb AAK33849.1
gi 13621632 gb AAK33423.1	gi 13622100 gb AAK33850.1
gi 13621635 gb AAK33426.1	gi 13622104 gb AAK33854.1
gi 13621643 gb AAK33433.1	gi 13622110 gb AAK33859.1
gi 13621655 gb AAK33444.1	gi 13622116 gb AAK33865.1
gi 13621656 gb AAK33445.1	gi 13622124 gb AAK33873.1
gi 13621659 gb AAK33448.1	gi 13622159 gb AAK33905.1
gi 13621673 gb AAK33461.1	gi 13622193 gb AAK33936.1
gi 13621686 gb AAK33473.1	gi 13622194 gb AAK33937.1
gi 13621696 gb AAK33482.1	gi 13622195 gb AAK33938.1
gi 13621703 gb AAK33488.1	gi 13622196 gb AAK33939.1
gi 13621712 gb AAK33497.1	gi 13622202 gb AAK33944.1
gi 13621728 gb AAK33511.1	gi 13622203 gb AAK33945.1
gi 13621738 gb AAK33520.1	gi 13622206 gb AAK33948.1
gi 13621739 gb AAK33521.1	gi 13622210 gb AAK33951.1
gi 13621740 gb AAK33522.1	gi 13622221 gb AAK33961.1
gi 13621772 gb AAK33551.1	gi 13622231 gb AAK33971.1
gi 13621776 gb AAK33555.1	gi 13622234 gb AAK33973.1
gi 13621791 gb AAK33569.1	gi 13622238 gb AAK33977.1
gi 13621798 gb AAK33575.1	gi 13622245 gb AAK33984.1
gi 13621801 gb AAK33578.1	gi 13622246 gb AAK33985.1
gi 13621803 gb AAK33580.1	gi 13622248 gb AAK33986.1
gi 13621804 gb AAK33581.1	gi 13622249 gb AAK33987.1
gi 13621832 gb AAK33606.1	gi 13622251 gb AAK33989.1
gi 13621833 gb AAK33607.1	gi 13622254 gb AAK33992.1
gi 13621896 gb AAK33665.1	gi 13622267 gb AAK34004.1
gi 13621897 gb AAK33666.1	gi 13622291 gb AAK34026.1
gi 13621906 gb AAK33674.1	gi 13622302 gb AAK34036.1
gi 13621911 gb AAK33679.1	gi 13622303 gb AAK34037.1
gi 13621949 gb AAK33713.1	gi 13622304 gb AAK34038.1
gi 13621951 gb AAK33715.1	gi 13622327 gb AAK34059.1
gi 13621962 gb AAK33724.1	gi 13622344 gb AAK34074.1
gi 13621963 gb AAK33725.1	gi 13622345 gb AAK34075.1
gi 13621964 gb AAK33726.1	gi 13622346 gb AAK34076.1
gi 13621971 gb AAK33732.1	gi 13622347 gb AAK34077.1
gi 13621976 gb AAK33737.1	gi 13622348 gb AAK34078.1
gi 13621983 gb AAK33744.1	gi 13622349 gb AAK34079.1

Tabl 34: List of GAS ORF's which ar shared with GBS but not with Spn

gi 13622382 gb AAK34109.1	gi 13622816 gb AAK34504.1
gi 13622386 gb AAK34113.1	gi 13622817 gb AAK34505.1
gi 13622391 gb AAK34118.1	gi 13622846 gb AAK34531.1
gi 13622392 gb AAK34119.1	gi 13622852 gb AAK34536.1
gi 13622397 gb AAK34123.1	gi 13622874 gb AAK34556.1
gi 13622404 gb AAK34129.1	gi 13622889 gb AAK34570.1
gi 13622412 gb AAK34136.1	gi 13622891 gb AAK34572.1
gi 13622413 gb AAK34137.1	gi 13622892 gb AAK34573.1
gi 13622414 gb AAK34138.1	gi 13622897 gb AAK34577.1
gi 13622418 gb AAK34142.1	gi 13622902 gb AAK34582.1
gi 13622430 gb AAK34152.1	gi 13622904 gb AAK34584.1
gi 13622446 gb AAK34167.1	gi 13622916 gb AAK34595.1
gi 13622449 gb AAK34169.1	gi 13622923 gb AAK34601.1
gi 13622453 gb AAK34173.1	gi 13622934 gb AAK34611.1
gi 13622470 gb AAK34188.1	gi 13622953 gb AAK34628.1
gi 13622487 gb AAK34204.1	gi 13622954 gb AAK34629.1
gi 13622490 gb AAK34206.1	gi 13622960 gb AAK34635.1
gi 13622502 gb AAK34217.1	gi 13622968 gb AAK34642.1
gi 13622503 gb AAK34218.1	gi 13622980 gb AAK34653.1
gi 13622514 gb AAK34228.1	gi 13622987 gb AAK34659.1
gi 13622528 gb AAK34241.1	gi 13623012 gb AAK34682.1
gi 13622540 gb AAK34252.1	gi 13623013 gb AAK34683.1
gi 13622541 gb AAK34253.1	gi 13623014 gb AAK34684.1
gi 13622544 gb AAK34255.1	gi 13623015 gb AAK34685.1
gi 13622545 gb AAK34256.1	gi 13623016 gb AAK34686.1
gi 13622546 gb AAK34257.1	gi 13623018 gb AAK34687.1
gi 13622547 gb AAK34258.1	gi 13623022 gb AAK34691.1
gi 13622548 gb AAK34259.1	gi 13623029 gb AAK34697.1
gi 13622550 gb AAK34261.1	gi 13623037 gb AAK34704.1
gi 13622551 gb AAK34262.1	gi 13623055 gb AAK34721.1
gi 13622552 gb AAK34263.1	gi 13623060 gb AAK34725.1
gi 13622556 gb AAK34267.1	gi 13623061 gb AAK34726.1
gi 13622557 gb AAK34268.1	gi 13623063 gb AAK34728.1
gi 13622558 gb AAK34269.1	gi 13623066 gb AAK34731.1
gi 13622559 gb AAK34270.1	gi 13623068 gb AAK34732.1
gi 13622563 gb AAK34273.1	gi 13623092 gb AAK34754.1
gi 13622571 gb AAK34281.1	gi 13623097 gb AAK34758.1
gi 13622576 gb AAK34286.1	gi 13623104 gb AAK34765.1
gi 13622581 gb AAK34290.1	gi 13623126 gb AAK34785.1
gi 13622582 gb AAK34291.1	gi 13623130 gb AAK34789.1
gi 13622586 gb AAK34295.1	gi 13623137 gb AAK34795.1
gi 13622589 gb AAK34298.1	gi 13623153 gb AAK34810.1
gi 13622605 gb AAK34312.1	gi 13623164 gb AAK34820.1
gi 13622633 gb AAK34338.1	gi 13623178 gb AAK34833.1
gi 13622635 gb AAK34340.1	
gi 13622637 gb AAK34342.1	
gi 13622638 gb AAK34343.1	
gi 13622657 gb AAK34360.1	
gi 13622707 gb AAK34404.1	
gi 13622716 gb AAK34413.1	
gi 13622724 gb AAK34420.1	
gi 13622732 gb AAK34427.1	
gi 13622743 gb AAK34437.1	
gi 13622761 gb AAK34453.1	
gi 13622773 gb AAK34464.1	
gi 13622788 gb AAK34478.1	

Table 35: GAS ORF's which are shared with pneumococcus but not with GBS

gi 13621338 gb AAK33157.1	gi 13623027 gb AAK34695.1
gi 13621352 gb AAK33168.1	gi 13623087 gb AAK34749.1
gi 13621410 gb AAK33221.1	gi 13623101 gb AAK34762.1
gi 13621433 gb AAK33242.1	gi 13623144 gb AAK34802.1
gi 13621445 gb AAK33253.1	gi 13623146 gb AAK34804.1
gi 13621446 gb AAK33254.1	gi 13623147 gb AAK34805.1
gi 13621447 gb AAK33255.1	
gi 13621448 gb AAK33256.1	
gi 13621449 gb AAK33257.1	
gi 13621451 gb AAK33259.1	
gi 13621460 gb AAK33267.1	
gi 13621466 gb AAK33272.1	
gi 13621489 gb AAK33293.1	
gi 13621490 gb AAK33294.1	
gi 13621519 gb AAK33320.1	
gi 13621520 gb AAK33321.1	
gi 13621653 gb AAK33443.1	
gi 13621722 gb AAK33506.1	
gi 13621723 gb AAK33507.1	
gi 13621724 gb AAK33508.1	
gi 13621805 gb AAK33582.1	
gi 13621900 gb AAK33669.1	
gi 13622011 gb AAK33769.1	
gi 13622212 gb AAK33953.1	
gi 13622280 gb AAK34016.1	
gi 13622381 gb AAK34108.1	
gi 13622409 gb AAK34134.1	
gi 13622410 gb AAK34135.1	
gi 13622423 gb AAK34146.1	
gi 13622428 gb AAK34151.1	
gi 13622441 gb AAK34162.1	
gi 13622442 gb AAK34163.1	
gi 13622454 gb AAK34174.1	
gi 13622456 gb AAK34176.1	
gi 13622619 gb AAK34325.1	
gi 13622642 gb AAK34346.1	
gi 13622643 gb AAK34347.1	
gi 13622664 gb AAK34366.1	
gi 13622666 gb AAK34368.1	
gi 13622667 gb AAK34369.1	
gi 13622671 gb AAK34372.1	
gi 13622672 gb AAK34373.1	
gi 13622673 gb AAK34374.1	
gi 13622674 gb AAK34375.1	
gi 13622679 gb AAK34380.1	
gi 13622680 gb AAK34381.1	
gi 13622682 gb AAK34382.1	
gi 13622755 gb AAK34448.1	
gi 13622758 gb AAK34450.1	
gi 13622759 gb AAK34451.1	
gi 13622835 gb AAK34521.1	
gi 13622837 gb AAK34523.1	
gi 13622937 gb AAK34614.1	
gi 13622942 gb AAK34618.1	
gi 13622946 gb AAK34622.1	
gi 13622978 gb AAK34652.1	

Table 36: Spn ORF's are shared with GBS and GAS

SP0001	SP0158	SP0254	SP0385
SP0002	SP0173	SP0259	SP0386
SP0003	SP0179	SP0261	SP0387
SP0004	SP0180	SP0262	SP0400
SP0005	SP0184	SP0263	SP0401
SP0006	SP0185	SP0264	SP0402
SP0007	SP0186	SP0265	SP0403
SP0008	SP0187	SP0266	SP0404
SP0010	SP0189	SP0268	SP0405
SP0011	SP0192	SP0271	SP0406
SP0013	SP0194	SP0272	SP0408
SP0014	SP0197	SP0273	SP0410
SP0019	SP0199	SP0274	SP0411
SP0021	SP0202	SP0280	SP0412
SP0024	SP0204	SP0281	SP0415
SP0027	SP0205	SP0282	SP0416
SP0032	SP0208	SP0283	SP0417
SP0033	SP0209	SP0284	SP0418
SP0034	SP0210	SP0285	SP0419
SP0035	SP0211	SP0286	SP0420
SP0036	SP0212	SP0287	SP0421
SP0037	SP0213	SP0289	SP0422
SP0042	SP0214	SP0290	SP0423
SP0044	SP0215	SP0291	SP0424
SP0045	SP0216	SP0292	SP0425
SP0046	SP0217	SP0294	SP0426
SP0047	SP0218	SP0295	SP0427
SP0048	SP0219	SP0303	SP0433
SP0051	SP0220	SP0310	SP0434
SP0053	SP0221	SP0314	SP0435
SP0054	SP0222	SP0317	SP0436
SP0056	SP0224	SP0318	SP0437
SP0063	SP0225	SP0319	SP0438
SP0073	SP0226	SP0320	SP0439
SP0074	SP0227	SP0321	SP0441
SP0078	SP0228	SP0322	SP0442
SP0079	SP0229	SP0323	SP0443
SP0083	SP0230	SP0324	SP0452
SP0084	SP0231	SP0325	SP0453
SP0085	SP0232	SP0327	SP0454
SP0095	SP0233	SP0330	SP0457
SP0105	SP0234	SP0334	SP0458
SP0106	SP0235	SP0336	SP0459
SP0111	SP0236	SP0337	SP0461
SP0112	SP0240	SP0338	SP0466
SP0118	SP0242	SP0340	SP0467
SP0120	SP0243	SP0342	SP0474
SP0121	SP0245	SP0369	SP0477
SP0122	SP0246	SP0370	SP0478
SP0127	SP0247	SP0371	SP0483
SP0128	SP0248	SP0373	SP0486
SP0129	SP0249	SP0374	SP0488
SP0148	SP0250	SP0381	SP0489
SP0149	SP0251	SP0382	SP0493
SP0151	SP0252	SP0383	SP0494
SP0152	SP0253	SP0384	SP0499

Tabl 36: Spn ORF's are shared with GBS and GAS

SP0500	SP0652	SP0787	SP0895
SP0501	SP0657	SP0788	SP0896
SP0502	SP0660	SP0792	SP0897
SP0515	SP0662	SP0793	SP0904
SP0516	SP0663	SP0797	SP0905
SP0517	SP0665	SP0798	SP0908
SP0519	SP0668	SP0799	SP0909
SP0521	SP0669	SP0801	SP0912
SP0522	SP0671	SP0802	SP0923
SP0523	SP0672	SP0803	SP0927
SP0526	SP0673	SP0805	SP0928
SP0549	SP0674	SP0806	SP0929
SP0550	SP0675	SP0807	SP0931
SP0552	SP0676	SP0816	SP0932
SP0553	SP0678	SP0817	SP0933
SP0554	SP0680	SP0820	SP0935
SP0555	SP0681	SP0822	SP0936
SP0556	SP0687	SP0823	SP0937
SP0557	SP0688	SP0824	SP0938
SP0563	SP0689	SP0825	SP0943
SP0567	SP0690	SP0828	SP0944
SP0568	SP0701	SP0829	SP0945
SP0576	SP0702	SP0831	SP0946
SP0577	SP0709	SP0835	SP0947
SP0578	SP0713	SP0837	SP0948
SP0579	SP0726	SP0838	SP0954
SP0581	SP0727	SP0839	SP0955
SP0588	SP0729	SP0841	SP0959
SP0589	SP0735	SP0843	SP0960
SP0591	SP0736	SP0844	SP0961
SP0592	SP0741	SP0845	SP0962
SP0593	SP0744	SP0846	SP0964
SP0603	SP0745	SP0847	SP0966
SP0604	SP0746	SP0848	SP0967
SP0605	SP0756	SP0851	SP0968
SP0608	SP0757	SP0852	SP0969
SP0610	SP0758	SP0855	SP0970
SP0611	SP0760	SP0856	SP0971
SP0613	SP0761	SP0862	SP0972
SP0614	SP0762	SP0864	SP0974
SP0615	SP0764	SP0865	SP0975
SP0616	SP0765	SP0867	SP0976
SP0618	SP0766	SP0868	SP0978
SP0620	SP0767	SP0869	SP0979
SP0622	SP0768	SP0870	SP0980
SP0623	SP0770	SP0871	SP0981
SP0624	SP0771	SP0872	SP0984
SP0626	SP0775	SP0873	SP0985
SP0630	SP0776	SP0875	SP0987
SP0631	SP0778	SP0876	SP0988
SP0636	SP0779	SP0877	SP0989
SP0637	SP0780	SP0878	SP0991
SP0638	SP0782	SP0880	SP0992
SP0645	SP0784	SP0881	SP0993
SP0646	SP0785	SP0893	SP1002
SP0647	SP0786	SP0894	SP1003

Table 36: Spn ORF's are shared with GBS and GAS

SP1004	SP1117	SP1242	SP1387
SP1008	SP1118	SP1244	SP1388
SP1010	SP1119	SP1245	SP1389
SP1012	SP1128	SP1246	SP1390
SP1016	SP1151	SP1247	SP1393
SP1017	SP1152	SP1248	SP1394
SP1018	SP1155	SP1249	SP1395
SP1020	SP1156	SP1260	SP1396
SP1021	SP1157	SP1263	SP1397
SP1022	SP1159	SP1266	SP1398
SP1024	SP1160	SP1275	SP1399
SP1025	SP1161	SP1276	SP1400
SP1026	SP1162	SP1277	SP1402
SP1029	SP1163	SP1278	SP1403
SP1033	SP1164	SP1279	SP1404
SP1034	SP1167	SP1280	SP1405
SP1035	SP1168	SP1283	SP1406
SP1045	SP1169	SP1284	SP1407
SP1056	SP1174	SP1285	SP1408
SP1067	SP1175	SP1286	SP1409
SP1068	SP1176	SP1287	SP1411
SP1069	SP1177	SP1288	SP1412
SP1070	SP1178	SP1289	SP1413
SP1071	SP1179	SP1290	SP1414
SP1072	SP1180	SP1291	SP1415
SP1073	SP1182	SP1293	SP1416
SP1074	SP1184	SP1297	SP1420
SP1076	SP1185	SP1298	SP1421
SP1079	SP1187	SP1299	SP1427
SP1081	SP1190	SP1308	SP1428
SP1082	SP1191	SP1316	SP1429
SP1083	SP1192	SP1324	SP1434
SP1084	SP1193	SP1329	SP1435
SP1087	SP1197	SP1330	SP1445
SP1088	SP1200	SP1331	SP1446
SP1089	SP1202	SP1336	SP1448
SP1090	SP1204	SP1341	SP1449
SP1093	SP1205	SP1354	SP1450
SP1094	SP1207	SP1355	SP1452
SP1095	SP1208	SP1357	SP1453
SP1096	SP1212	SP1358	SP1456
SP1097	SP1213	SP1359	SP1457
SP1098	SP1218	SP1362	SP1458
SP1099	SP1219	SP1368	SP1460
SP1100	SP1220	SP1370	SP1461
SP1102	SP1225	SP1371	SP1462
SP1105	SP1226	SP1372	SP1465
SP1106	SP1227	SP1374	SP1466
SP1107	SP1228	SP1375	SP1469
SP1110	SP1229	SP1376	SP1470
SP1111	SP1230	SP1377	SP1473
SP1112	SP1231	SP1378	SP1474
SP1113	SP1232	SP1380	SP1475
SP1114	SP1233	SP1381	SP1478
SP1115	SP1238	SP1383	SP1479
SP1116	SP1241	SP1386	SP1482

Table 36: Spn ORF's are shared with GBS and GAS

SP1483	SP1580	SP1685	SP1857
SP1485	SP1583	SP1688	SP1858
SP1489	SP1584	SP1689	SP1860
SP1491	SP1586	SP1697	SP1861
SP1498	SP1587	SP1698	SP1865
SP1500	SP1588	SP1699	SP1871
SP1501	SP1589	SP1702	SP1873
SP1502	SP1590	SP1709	SP1874
SP1504	SP1591	SP1711	SP1875
SP1505	SP1597	SP1712	SP1876
SP1507	SP1598	SP1713	SP1877
SP1508	SP1599	SP1714	SP1878
SP1509	SP1602	SP1717	SP1879
SP1510	SP1603	SP1721	SP1880
SP1511	SP1606	SP1722	SP1881
SP1512	SP1608	SP1724	SP1883
SP1513	SP1609	SP1725	SP1884
SP1517	SP1610	SP1726	SP1887
SP1518	SP1615	SP1727	SP1888
SP1519	SP1616	SP1732	SP1889
SP1521	SP1617	SP1733	SP1890
SP1522	SP1624	SP1734	SP1895
SP1523	SP1625	SP1735	SP1896
SP1529	SP1626	SP1736	SP1900
SP1530	SP1631	SP1737	SP1901
SP1534	SP1633	SP1738	SP1902
SP1535	SP1638	SP1739	SP1903
SP1536	SP1644	SP1742	SP1906
SP1537	SP1645	SP1743	SP1908
SP1538	SP1646	SP1744	SP1909
SP1539	SP1647	SP1746	SP1916
SP1540	SP1648	SP1747	SP1918
SP1541	SP1649	SP1748	SP1922
SP1542	SP1650	SP1749	SP1940
SP1544	SP1652	SP1750	SP1942
SP1547	SP1653	SP1752	SP1944
SP1549	SP1655	SP1759	SP1953
SP1551	SP1659	SP1776	SP1957
SP1552	SP1661	SP1780	SP1960
SP1553	SP1662	SP1781	SP1961
SP1554	SP1664	SP1782	SP1963
SP1557	SP1665	SP1785	SP1964
SP1558	SP1666	SP1790	SP1966
SP1559	SP1667	SP1795	SP1967
SP1560	SP1668	SP1799	SP1968
SP1561	SP1670	SP1804	SP1969
SP1563	SP1671	SP1816	SP1970
SP1564	SP1672	SP1817	SP1972
SP1565	SP1674	SP1825	SP1973
SP1566	SP1675	SP1839	SP1974
SP1568	SP1676	SP1840	SP1975
SP1569	SP1677	SP1845	SP1976
SP1571	SP1681	SP1847	SP1979
SP1574	SP1682	SP1848	SP1980
SP1575	SP1683	SP1851	SP1981
SP1577	SP1684	SP1855	SP1982

Table 36: Spn ORF's ar shared with GBS and GAS

SP1983	SP2085	SP2206
SP1984	SP2086	SP2207
SP1985	SP2087	SP2208
SP1987	SP2088	SP2209
SP1989	SP2090	SP2210
SP1990	SP2091	SP2214
SP1991	SP2092	SP2215
SP1993	SP2094	SP2216
SP1994	SP2099	SP2219
SP1996	SP2100	SP2220
SP1997	SP2101	SP2221
SP1998	SP2106	SP2222
SP1999	SP2107	SP2224
SP2006	SP2108	SP2225
SP2007	SP2109	SP2226
SP2010	SP2110	SP2227
SP2011	SP2112	SP2228
SP2012	SP2113	SP2229
SP2020	SP2114	SP2230
SP2021	SP2119	SP2231
SP2022	SP2121	SP2233
SP2027	SP2129	SP2234
SP2028	SP2131	SP2235
SP2030	SP2135	SP2238
SP2031	SP2142	SP2239
SP2032	SP2148	SP2240
SP2033	SP2150	
SP2034	SP2151	
SP2035	SP2152	
SP2036	SP2153	
SP2037	SP2156	
SP2038	SP2161	
SP2040	SP2162	
SP2041	SP2169	
SP2042	SP2170	
SP2044	SP2171	
SP2045	SP2172	
SP2048	SP2173	
SP2052	SP2174	
SP2053	SP2175	
SP2054	SP2176	
SP2055	SP2184	
SP2056	SP2185	
SP2057	SP2186	
SP2058	SP2187	
SP2063	SP2188	
SP2065	SP2189	
SP2069	SP2191	
SP2070	SP2192	
SP2072	SP2193	
SP2073	SP2194	
SP2075	SP2195	
SP2077	SP2202	
SP2078	SP2203	
SP2082	SP2204	
SP2083	SP2205	

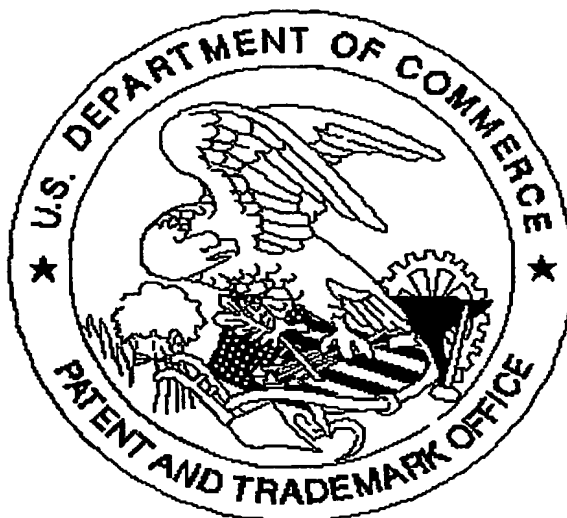
Tabl 37: Spn ORF's which are shared with GBS but not with GAS

SP0012	SP0725	SP1360	SP1927
SP0020	SP0730	SP1361	SP1928
SP0039	SP0739	SP1365	SP1943
SP0050	SP0749	SP1382	SP1959
SP0082	SP0750	SP1384	SP2001
SP0107	SP0751	SP1392	SP2002
SP0113	SP0752	SP1447	SP2009
SP0119	SP0753	SP1451	SP2026
SP0146	SP0754	SP1463	SP2029
SP0150	SP0769	SP1464	SP2039
SP0175	SP0789	SP1471	SP2061
SP0176	SP0791	SP1472	SP2064
SP0177	SP0826	SP1524	SP2066
SP0178	SP0900	SP1527	SP2079
SP0237	SP0913	SP1600	SP2084
SP0255	SP0914	SP1605	SP2095
SP0260	SP0939	SP1607	SP2096
SP0267	SP0941	SP1632	SP2098
SP0278	SP0942	SP1634	SP2103
SP0288	SP0953	SP1651	SP2127
SP0346	SP0973	SP1673	SP2128
SP0347	SP0977	SP1680	SP2130
SP0348	SP1011	SP1695	SP2134
SP0349	SP1013	SP1700	SP2137
SP0366	SP1027	SP1701	SP2138
SP0376	SP1054	SP1720	SP2157
SP0413	SP1055	SP1729	SP2196
SP0445	SP1080	SP1740	
SP0462	SP1086	SP1741	
SP0463	SP1121	SP1745	
SP0479	SP1122	SP1751	
SP0480	SP1123	SP1757	
SP0482	SP1124	SP1758	
SP0484	SP1126	SP1761	
SP0537	SP1127	SP1762	
SP0538	SP1137	SP1763	
SP0566	SP1166	SP1764	
SP0580	SP1173	SP1765	
SP0585	SP1194	SP1766	
SP0599	SP1195	SP1767	
SP0600	SP1215	SP1768	
SP0601	SP1240	SP1770	
SP0606	SP1256	SP1771	
SP0607	SP1261	SP1772	
SP0609	SP1271	SP1783	
SP0617	SP1272	SP1802	
SP0627	SP1273	SP1828	
SP0655	SP1274	SP1856	
SP0656	SP1306	SP1867	
SP0710	SP1310	SP1869	
SP0711	SP1332	SP1870	
SP0717	SP1333	SP1872	
SP0718	SP1334	SP1891	
SP0720	SP1346	SP1907	
SP0723	SP1348	SP1910	
SP0724	SP1350	SP1911	

Table 38: Spn ORF's which are shared with GAS but not with GBS

SP0065	SP1754
SP0075	SP1797
SP0090	SP1798
SP0091	SP1800
SP0092	SP1885
SP0099	SP1919
SP0100	SP1923
SP0153	SP1941
SP0155	SP1950
SP0156	SP2016
SP0200	SP2017
SP0306	SP2051
SP0313	SP2060
SP0341	SP2111
SP0476	SP2143
SP0496	SP2144
SP0509	SP2201
SP0527	SP2236
SP0648	
SP0658	
SP0659	
SP0661	
SP0677	
SP0715	
SP0742	
SP0743	
SP0858	
SP0859	
SP0860	
SP0910	
SP0986	
SP0994	
SP0999	
SP1000	
SP1001	
SP1023	
SP1075	
SP1129	
SP1147	
SP1171	
SP1186	
SP1315	
SP1317	
SP1319	
SP1320	
SP1321	
SP1322	
SP1438	
SP1442	
SP1525	
SP1546	
SP1570	
SP1572	
SP1578	
SP1604	
SP1715	

United States Patent & Trademark Office
Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☐ Page(s) _____ of _____ were not present
for scanning. (Document title)

☐ Page(s) _____ of _____ were not present
for scanning. (Document title)

✓ *Scanned copy is best available. Drawings fig. 1, 4 and 5
are very dark.*